


```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HK368 / SEROTYPE B;
RX MEDLINE=89379374; PubMed=2506130;
RA Poulsen K., Brandt J., Hjorth J.P., Thøgersen H.C., Killian M.;
RT "Cloning and sequencing of the immunoglobulin A1 protease gene (iga)
RL of Haemophilus influenzae serotype b."
RN Infect. Immun. 57:3097-3105(1989).
[2]
RP MUTAGENESIS OF SER-288.
RC STRAIN-HK368 / SEROTYPE B;
RX MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Killian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
RL influenzae type 1 immunoglobulin A1 proteases."
CC J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTRACYTIC FC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY). BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC -1- SIMILARITY:
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X64357; CAA45708.1; -
CC EMBL: M67492; AAA24969.1; -
CC DR MEROPS: S06.001; -
CC DR InterPro: IPR000710; IGA_S6.
CC DR Pfam: PF02395; IGA1.1.
CC DR PRINTS: PRO0921; IGASERPRASE.
CC KM Hydrolyase; Serine protease; Transmembrane; Zymogen; Signal.
CC FT STGNL 1 25 POTENTIAL.
CC FT CHAIN 26 1008 IMMUNOGLOBULIN A1 PROTEASE.
CC FT PROPEP 1009 1541 HELPER PEPTIDE (POTENTIAL).
CC FT ACT_SITE 288 288 PROBABLE.
CC FT MUTAGEN 288 288 S->T: LOSS OF ACTIVITY.
CC SQ SEQUENCE 1541 AA; 169370 MW; CE7257CB3196C600 CRC64;

alignment_scores:
Quality: 1608.50 Length: 1717
Ratio: 1.733 Gaps: 51
Percent Similarity: 54.048 Percent Identity: 27.257

alignment_block:
US-09-303-518D-649 x IGA1_HAEIN ..

Align seg 1/1 to: IGA1_HAEIN from: 1 to: 1541

64 CGCTTCGCGCTGCTTACTTACCATATGCGTGGTTCGGCATTCCTTC 113
   :::::::::::::::::::::
5 LyspHeuLysLeuAsnPhelLeuAlaLeuThrValAlaTyrAlaLeuThr 21
   :::::::::::::::::::::
114 CCAAGCGCTGGCGGACACACTTATTGGCATCAACACCAATACATATC 163
   :::::::::::::::::::::
21 CTTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38
   :::::::::::::::::::::
164 GGGAGCTTCCGCAAAATAAGCAAGTTGCGAGTCGGGCGCAAGATATT 213
   :::::::::::::::::::::
38 rGaSpHeuAlaGlnLysnLysGlyLysPheSerValGlyAlaThrAsnVal 54

```

```

214 GAGGTTTACACAAAAAGGGAGTGTGCGCAATCAATGACAAAGC 263
   ||| :::::::::::::::::::::
55 LeuValLysAspLysAsnAsnLysAspLeuLysThrAlaLeuProAsnGly 71
   ||| :::::::::::::::::::::
264 C...CCGATGATGATATTTCTGTGTGTGCG...CGTACGGCGTGGCGG 307
   ||| :::::::::::::::::::::
71 yLLeuProkettLeuAspPheSerValAlaAspValAspLysArgIleAla 88
   ||| :::::::::::::::::::::
308 CATGTGTGGCGGACATATATGTGTGAGCGGCGACAT..... 345
   :::::::::::::::::::::
88 hLeuLleAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104
   :::::::::::::::::::::
346 ..... AACGGCGGTATTAACAACGTGA 368
   ||| :::::::::::::::::::::
105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121
   ||| :::::::::::::::::::::
369 TTTTGGTGGCGGAGGAAATCCCGATCAACATCGTTTACTATATAAA 418
   :::::::::::::::::::::
121 nAlaLysAlaHis...ArgAspValSerSerGlnGlnLysnArgTyrPheS 137
   ||| :::::::::::::::::::::
419 TTGTGAACGGAATATTATTAAGCAGGAGCTAAAGC..... 456
   ||| :::::::::::::::::::::
137 eValGlnLysAsnGlnLysProThrLysLeuAsnGlyLysThrValThr 153
   ||| :::::::::::::::::::::
457 ..... CATCGTTATGGCGGCGATTTATCATATGCGCGGCTT 491
   ||| :::::::::::::::::::::
154 ThrGluAspGlnThrClnLysArgArgGlnAspTyrTyrMetProAlaLe 170
   :::::::::::::::::::::
492 GCATGAATTTGTACAGATGCGAGACCTGTGAA...ATGCACTATTATA 538
   ||| :::::::::::::::::::::
170 uAspLysPheValThrGlnValAlaProLleGlnLysSerThrAlaSerS 187
   ||| :::::::::::::::::::::
539 TGGATGGCGGAAATATATGATCAATAATATTACCGTGCACCGCTTGTG 588
   ||| :::::::::::::::::::::
187 eAspArgLagLysThrTyrAsnAspGlnAsnLysTyrProAlaPheValArg 203
   ||| :::::::::::::::::::::
589 ATTTGGGCGGAGCGAATATGGCGATGCTGATGAAGATGAG..... 630
   :::::::::::::::::::::
204 LeuGlySerGlySerGlnPheLleTyrLysLysGlyAspAsnTyrSerLe 220
   ||| :::::::::::::::::::::
631 .....CCCAATTAACCGGAAAGTTCAATATCAT.....ATTGCA 664
   ||| :::::::::::::::::::::
220 uLleLeuAsnAsnHisGlnValGlnLysAsnLysLeuLysValGlyAla 237
   ||| :::::::::::::::::::::
665 GTGCGTATTTCTGGCTGCTGGTGGCAATACCTTTGCAAAATGAGATCA 714
   ||| :::::::::::::::::::::
237 sPalatyrThrTyrGlyLleAlaGlyThrProTyrLysValAsnHisGln 253
   ||| :::::::::::::::::::::
715 GGTGGTGGCGACAGTCAACTAGTAGTGAATAATTAACATPAGC...CC 761
   :::::::::::::::::::::
254 AsnAsnGlyLeuLleGlyPheGlyAsnSerLysGlnGlnHisSerAspTr 270
   ||| :::::::::::::::::::::
762 ATATGTTTTTTTACCAACAGGA.....GGCTCATTTGGCG 796
   ||| :::::::::::::::::::::
270 oLysGlyLleLeuSerGlnAspProLeuThrAsnTyrAlaValLeuGlyA 287
   ||| :::::::::::::::::::::
797 ACAAGTGGCTACCAAGTTTATCTATGATGCGCCAAAGCAAGAGGTGTA 846
   ||| :::::::::::::::::::::
287 sPserGlySerProLeuPheValTyrAspArgGlnLysGlyStrPheLeu 303
   ||| :::::::::::::::::::::
847 ATTAATGGGATATTGCAACGCGCAACCCCTATATAGAAAGCAATGG 896
   :::::::::::::::::::::
304 PheLeuGlySerTyrArgPheTrpAlaGlyTyrAsnLysLysSer..... 318
   ||| :::::::::::::::::::::
897 CTTCCAGCTGGTTCGTAAGATTTGG...TTCTATGATGAATTTTGGCTG 943
   :::::::::::::::::::::
319 .....TrpGlnLutTrpAsnLleTyrLysSerGlnPheThrL 331
   ||| :::::::::::::::::::::
944 GAGATACCATTACGATTTCTACGAACCCAGCTCAAAATGGAAATACTCT 993
   ||| :::::::::::::::::::::
331 yAspVal..... 333

```

[illegible]

```
851 ..... 851
2601 TTTGAAGACAGCCGCTTACGGACAAATCAGCGCGGCAAGATACGG 2650
851 ..... 851
2651 CATTACACTTAAAAAGACAGCAATGACGCTCCGTCAGGACGAAATTA 2700
852 ..... 852
852 ..... SerHisTrpHisLeuThrGlnLysSerAspVal 862
2701 GGCAATTAAACCTTGACACAGCCACCATTAACATTCGCGCTATCG 2750
863 HisGlnLeuAspLeuAlaAsnGlnHisIleHisLeuAsnSerAlaAsp 879
2751 CCACGATGCGGCAAGGGCGCAACCGGCACTGCGACAGATGCGCGCC 2800
879 nSerAsn..... 881
2801 GCCGTTGCGCGCGTTCGCGCGCTCCCTATTCGCTTACACCGCACT 2850
882 ..... AsnValThr..... 884
2851 TCGGTAGATCCGTTTCAACAGCCTGACGCTAAACGGCAATTTGAACG 2900
885 ..... LysTyrAsnThrLeuThrValAsn...SerLeuSerGln 896
2901 TCACGGACATTCGCGCTTTATGTCGGAACCTTCGCTACCGGACGACA 2950
896 yAsnGlySerPheTyrLeuThrAspLeuSerAsnLysGlnGlyAsp 913
2951 AATTGAAGCGGGGCAAGTTTCCGAGGCACTTACACCTTGCGGTCAC 3000
913 yValValValThrLysSerAlaThrGlnLysnThrLeuGlnValAla 929
3001 AATACCGCAACGAACTCGAAGCCTCGAACAATTTGACGTAAGTAGG 3050
930 AspLysThrGlnGlyPro...AsnHisAsnIleLeuThrLeuPheAsp 945
3051 AAAAGACAAACCGCTGTCCGAAACCTTAATTCACCGCTGCAAAACG 3100
945 a.....SerLysAlaGlnArgAspHisLeuAsnValSerLeuValGly 960
3101 AACAGTCGATGCGCGCGCTGACGCTTACCAACATTCGCGCAAGACGCG 3150
960 snThrValAspLeuGlnValLeuThrLysLysLeuArgAsnValAsnGly 976
3151 GAGTTCGCGCTGACATATCCGCTCAAGACAA..... 3183
977 ArgTyrAspLeuTyrAsnProGlnValGlnLysArgAsnGlnThrVal 993
3183 ..... 3183
993 pThrThrAsnIleThrThrProAsnAsnIleGlnAlaAspValProSer 1010
3184 ..... GAGCTTTCGCAAACTCGGCAAG..... 3207
1010 alProSerAsnAsnGlnGlnIleAlaArgValAspGlnAlaProValPro 1026
3208 ..... GCAGAGC 3215
1027 ProProAlaProAlaThrProSerGlnThrThrGlnThrValAlaGln 1043
3216 CAAAACAGCGCGGAAAA.....GACAAACCGCAAAAGCCTTGACG 3256
1043 nSerLysGlnLysThrValGlnLysAsnGlnGlnAspAlaThrG 1060
3257 CGCTGATTCGCGCGCGCGCGATTCGCGCA..... 3288
1060 lThrThrAlaGlnAsnArgGlnValAlaLysGlnAlaLysSerAsnVal 1076
3289 AAGACAGAAAGCGTTGCGCAACCGCGCGGACGAGCGAGGGAATGT 3338
1077 LysAlaAsnThrGlnThrAsnGlnValAlaGlnSerGlnThrLys 1093
3339 CGCATTTATGACGCGAG.....GAG 3361
1093 sGlnThrGlnThrThrGlnThrLysGlnThrAlaThrValGlnLysGln 1110
3362 AGAAAAAGCGGTGACGCGGATTAAGACACCGCTTGCGGAA..... 3405
1110 lLysAlaLysValGlnThrGlnLysThrGlnGlnValProLysValThr 1126
3406 ..... CAGCCGAGAC 3416
1127 SerGlnValSerProLysGlnGlnLysSerGlnThrValGlnProGlnAl 1143
3417 GGA..... 3420
1143 aGlnProAlaArgGlnAsnAspProThrValAsnIleLysGlnProGln 1160
3420 ..... 3420
1160 ergLInThrAsnThrThrAlaAspThrGlnGlnProAlaLysGlnThrSer 1176
3420 ..... 3420
1177 SerAsnValGlnGlnProValThrGlnSerThrThrValAsnThrGlnLys 1193
3421 ..... ACCCGCGCGCTACACCGCGCTTC 3445
1193 nSerValValGlnAsnProGlnLysnThrThrProAlaThrGln...P 1209
3446 CC.....CGCGCGCGCG 3459
1209 rThrValAsnSerGlnSerSerAsnLysProLysAsnHisArgArg 1225
3460 GCCCGCGGATTTGCG...CACTGCACCCCAACCGCACCGCCCAAC 3506
1226 SerValArgSerValProHisAsnValGlnProAlaThrThrSerSer 1242
3507 GCAGCGC.....GACCTGATCAGCCGTTAGCCATA 3538
1242 nAspArgSerThrValAlaLeuLysAspLeuThrSerThrAsnThrAsn 1259
3539 CGCGTTTGATGATTTTCGCC.....ACGCTCAACG 3573
1259 lValLeuSerAspAlaArgAlaLysAlaGlnPheValAlaLeuAsnVal 1275
3574 GTTTCGCGTACAGGACGAATTAGCGCGTA...TTGCCGAAGACCG 3620
1276 GlyValAlaValSerGlnHisIleSerGlnLeuGlnMetAsnAsnGln 1292
3621 CGCGCAACGCGCTTGACAAAGCGGCTCGGACACCAACACTACCGTT 3670
1292 yGlnTyrAsnValTyrValSerAsnThrSerMetAsnLysAsnTyrSer 1309
3671 CGCAATTTCCGCGCTACCGCGCAACAAACGACTGCGCAATCGT 3720
1309 eSerGlnTyrArgArgPheSerSerLysSerThrGlnThrGlnLeuGly 1325
3721 ATCGAGAAAACTCGCGACGCGG...CGCGTGCATCTGTTTCGA 3767
1326 TrpAspLInThrIleSerAsnAsnValGlnLeuGlnGlyValPheThrTy 1342
3768 CAACCGGACCGAACAACCTTCGACGACGCAACGCACTGCGCAATCGC 3817
1342 rValArgAsnSerAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsn 1358
3818 TTGCGCACGCGCGCTTTCGCAATACGCAATCGAC...ACGTTCTAC 3864
1358 eValGlnValAsnPheTyrSerLysTyrTyrAlaAspAsnHisTyrTy 1374
3865 ATCGCGATTCAGCGCGCGCGGCTTTAGACAGGCGGACGCTTTCAGACG 3914
1375 LeuLysIleAspLeuGlnTyrGlyLysPheGlnSerLysLeuGlnThr 1391
```


2000 GGGAAATGCTGGGACACAGACTGTGATCAACCGCACATTTAAAGCGAA 2049
732 snglvalvalvalgluaspasprtrlleasnargspnphelysalathr 748
2050 AACTTCCAAATTAAGCGGACGGGGGTGGCTTCC... CGCAAGCTTGC 2096
749 AsnIleasnvalThrAsnAsnAlaThrLeuTyrserylArgAsnValGI 765
2097 CAAAGTCMAAGCGGATTCGACATTTGACAAATCAGCCCAAGCAGTTTTCG 2146
765 uSerIleThrSerAsnIleThrAlaSerAsnAlaLysValHisIleGG 782
2147 GTGTGCGACCGCATCAAAAGCCACAGCAATCTGATACAGTTCCGACTGGACG 2196
782 LylTr... LysAlaGlyAspThrValLysValAlaIgsersAspTyrThr 796
2197 GGTGTGCAAAATTTGTGTGCAAAAACCATTCAGGACGATTAAGTATTCG 2246
797 GlyTyrValThrCysThrThrAspLysLeuSer... AspLysAlaLeuAs 812
2247 TTCAATTGACTAAGCGACATCAGCGGCGCATGTGCATCTTCCGATCAGC 2296
812 nSerPheAsnProThrAsnLeuArgGlyAsnValAsnLeuThrGluSerA 829
2297 CTCATTTAAATCTCACAGGGCTTGCACACTCAACGGCATCTTATGTGCA 2346
829 IAsnPheValLeu... GlyLysAlaAsnLeuPheGlyThrIleGlnSer 844
2347 AATGGCGATACAGCTTATATACACTGACGCCACAAAGCCCAACAAAGCGCAA 2396
845 ArgGlyAsnSerGluValAlaArgLeuThrGluAsn... 855
2397 CCTTAGCCTCTGTGGCAATGCCCAAGCAACATTTAATCAAGCCACATTA 2446
855 855
2447 ACGGCACACATCGGCTTCGGGCAATGCTTCATTTAATCAAGCCACAC 2496
855 855
2497 GCCGTCAAAAACGGAGCTGTACGCTTCGGGCAACCGCTAAGCAAACT 2546
855 855
2547 AAGCCATTCGCGACTCAAGGTAATGTCTCCCTAGCCGATTAAGCAGATAT 2596
855 855
2597 TCCATTTTAAAGCAGCCGCTTTACCGGACAAATCAGCGCGCGCAAGGAT 2646
855 855
2647 ACGGCATTACACTTAAAGACAGGAAATGACGCTGCCGTGACGACGGA 2696
856 SerHisTrpHisLeuThrGlyLysnSerAs 865
2697 ATTAGCAATTTAAACCTTGACAAAGCCACCATTAACATTCATATTCGGCT 2746
865 PAlaHisGlnLeuAspLeuAlaAsnGlyHisIleHisLeuAsnSerAlaA 882
2747 ATCGCCACGATATCGGACAGGCGCAACCGGCAAGTGGACAGATGGCGCG 2796
882 sPAsnSerAsn... 885
2797 CGCGCCGCTTCGGCGCGCTTCGCCGCTTCCTATATCCGTTACACCGCC 2846
886 AsnValTr... 888
2847 AACTTCGGTAGAATCCGTTTCAACAGCGTACGGTAACGCGCAATTCGA 2896
889 LysTyrAsnThrLeuThrValAsn... SerLeuAs 899

```

3421 .....ACCGGCGCGGTACACCGCC 3441
1196 LysSerValValGluAsnProGluAsnThrThrProAlaThrGln 1212
3442 TTCGCC.....CGGCGCG 3455
1213 ...ProThrValAsnSerGluSerSerAsnLysProLysAsnArgHisArg 1228
3456 CGCGCGCGCGCGGATTTCCG...CAATGCAACCCCAACCGCGCGCC 3502
1228 GAGSerValArgSerValProHisAsnValGluProAlaThrThrSerS 1245
3503 AACCGCAGCGC.....GACCTGATCAGCCGTATGCC 3534
1245 GAsnAspArgSerThrValAlaLeuGlyAspLeuThrThrAsnThr 1261
3535 AATAGCGGTTGGTGAATTTCCGCC.....ACGCTCAA 3569
1262 AsnAlaValLeuSerAspAlaAlaValAlaGlnPheValAlaLeuS 1278
3570 CAGCGTTTTCGCGTACAGCAGCAATTAAGCGCGTA...TTTGCCTGAAG 3616
1278 ValGlyLysAlaValSerGlnHisIleSerGlnLeuGluMetAsnArg 1295
3617 ACCGCGCGCAACCGCGTTTGACAAAGCGGCATCCGGGACACCAACACTAC 3666
1295 LuGlyGlnThrAsnValThrValSerAsnThrSerMetAsnLysAsnThr 1311
3667 CGTTCCGCAATTTCCGCGCTACCGCCCAACAAACGACCTGCGCAAT 3716
1312 SerSerSerGlnThrArgArgPheSerSerLysSerThrGlnThrGln 1328
3717 CGGTATGCAGAAAACCTCGCAGCGGG...CGCGTGGCATCCTGTTT 3763
1328 uGlyThrAspGlnThrIleSerAsnAsnValGlnLeuGlyValAlaPheT 1345
3764 CGCAGCAGCGGACCGGAAACACTCTGACGAGCGCATCGCACTCGGCA 3813
1345 hTrpValAlaArgAsnSerAsnAsnAsnAspLysAlaThrSerLysAsnThr 1361
3814 CGGCTTGCGCCAGCGCGCTTTTGGGCATACGGCATCGAC...AGGTT 3860
1362 ..LeuAlaGlnValAsnPheThrLysThrLysAlaAspAsnHisThr 1377
3861 CTATAGCGGCATCAGCGCGCGGCTTTTACAGCGCGAGCTTTCAG 3910
1377 PTrpLeuGlyIleAspLeuGlyTrpGlyLysPheGlnSerLysLeuGlnT 1394
3911 ACGGATCGAGGCAAAATCCGCGCGCGCTGTGATTAACGAGATTACG 3960
1394 hArgHisAsnAlaLysPheAlaArgHisThrAlaGlnPheGlyLeuThr 1410
3961 GCAGCATACCGCGCGGTTTGGCGGATTCGGCATCGCAACCGCAGTCG 4010
1411 AlaGlyLysAlaPheAsnLeuGlyAsnPheGlyIleThrProIleValG 1427
4011 CGCAGCGCGCTATTTCGTCAAAAGCGATTACCGCTACGAAAACGTCA 4060
1427 yValArgLysSerTrpLeuSerAsnAlaAspPheAlaLeuAspGlnAlaA 1444
4061 AATATCGCAACCCCGCGCTTGCAATTAACCGCTACCGCGCGCATTAAG 4110
1444 gLleLysValAsnProIleSerValLysThrAlaPheAlaGlnValAsp 1460
4111 GAGATTTATTCATTAACCGCGCGCAACATTCATTAACCGCTTATTT 4160
1461 LeuSerTrpThrTrpHisLeuGlyLys...PheSerValThrProIleLe 1476
4161 GAGCTGTTCCTATACGATCGCGCTTGGGCAAAAGTCCGAACCGCTCA 4210
1476 uSerAlaAlaGlyTr...AspAlaAsnGlnGlySerGlyLysIleAsnValA 1492
4211 AATACCGCGCTATTGGCTCAGATTTCCGCAAAACCGCGCATGCGAATG 4260

```

```

1492 snGlyTrpAspPheAlaTrpAsnValGluAsnGlnGlnGlnTrpAsnAla 1508
4261 GCGCTAAACGCCGAATCAAAAGCTTTTACCGCTGTCCCTCCACCGCTGCCG 4310
1509 GlyLeuLysLeuLysTrpHisAsnValLysLeuSerLeuLeuGlyGlyLe 1525
4311 CGCCAAAGCGCGCACTGGAAGCAACACAGCGCGCGCATCAAAATTAG 4360
1525 uThrLysAlaLysGlnAlaGlyLysGlnLysThrAlaGluLeuLysLeuS 1542
4361 GCTAC 4365
1542 erPhe 1543

seq_name: SwissProt_40:IGA_NEIGO

seq_documentation_block:
ID IGA_NEIGO STANDARD; PRT; 1532 AA.
AC P09790;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE IGA-specific serine endopeptidase precursor (EC 3.4.21.72) (IGA
DE protease).
DE IGA.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID:485;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC STRAIN=MS11;
RA MEDLINE-87115823; PubMed-3027577;
RT Pohlner J., Halter R., Beyreuther K., Meyer T.F.;
RT "Gene structure and extracellular secretion of Neisseria gonorrhoeae
RT IGA protease."
RT Nature 325:458-462(1987).
RL [2]
RN ACTIVE SITE.
RP MEDLINE-90154052; PubMed-2105953;
RA Bachovchin W.W., Plant A.G., Flentke G.R., Lynch M., Kettner C.A.;
RT "Inhibition of IGA1 proteinases from Neisseria gonorrhoeae and
RT Hemophilus influenzae by peptide prolyl boronic acids."
RL J. Biol. Chem. 265:3738-3743(1990).
CC -1- FUNCTION: THIS PROTEASE IS SPECIFIC FOR IMMUNOGLOBULIN A.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDE THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL, X04835; CAZ28538.1; -.
DR PIR, A26039; A26039.
DR MEROPS, S06.001; -.
DR InterPro, IPR000710; IGA_S6.
DR Pfam, PF02395; IGA1; 1.
DR PRINTS, PR00921; IGASERPTASE.
KW Hydrolase; Serine protease; zymogen; Autocatalytic cleavage;
KW Transmembrane; Signal.
FT SIGNAL 1 27
FT CHAIN 28 986 IGA-SPECIFIC SERINE ENDOPEPTIDASE.

```

468 pLysLysValGlnAlaPheSerGlnValGlyIleValSerGlyArgGlyT 485


```

949 laAsnHisTyrValAspLeuGluValLeuArgTyrThrIleLeuThrGlu 965
3145 GAGCGCGAGTTCGCGCTCATAT.....CGGCT 3173
3174 CAAGAACAGAGCTTTCGAAACCTCGCAGACAGAACCCAAAAAC 3223
982 llyserProAlaProSerProAlaAlaAsnThrAlaSerGlnAlaGlnLysa 999
3224 AGCGGGAAGAAAGACAGACCGCAAGCCCTTCAGCGCTGATTCGCGCGG 3273
999 laThrGlnThrAspGlyAlaGln.....IleAlaLysPro 1010
3274 CGCGATCGCTCGAAAGACAGAAAGCGTTCGACGCGGCG..... 3315
1011 GlnAsnIleVal.....ValAlaProProSerProGlnAl 1022
3316 .....CGCAGCAGCGCGGGAATGTCG 3340
1022 AAsnGlnAlaGlnGluAlaLeuAlaArgGlnAlaLysAlaGlnValL 1039
3341 GCATTATCCAGCG.....GAGGA 3360
1039 ysArgGlnGlnAlaAlaGlnAlaGlnLysValAlaArgGlnLysAspGln 1055
3361 GAGAAAAACGGGTGACGCGGAT..... 3384
1056 GlnAlaLysArgLysAlaAlaGlnIleAlaArgGlnGlnGlnAlaArg 1072
3385 .....AAG 3388
1072 gLysAlaAlaGlnLeuAlaAlaLysGlnLysAlaGlnAlaGlnArgLysa 1089
3389 ACACCGCTTGGCAACAGCGCAGACCGCAACCGCGCGCTACACAC 3438
1089 laArgGlnLeuAlaArgGlnLysAlaGlnLysAlaSerHisGlnAlaAsn 1105
3439 GCCTTCCCGCGCGCGCGCGCGCGCGGATTCGCGCA..... 3480
1106 AlaLysProLysArgArgArgArgAlaIleLeuProArgProAl 1122
3480 ..... 3480
1122 aproValPheSerLeuAspAspTyrAspAlaLysAspAsnSerGlnSer 1139
3480 ..... 3480
1139 erIleGlyAsnLeuAlaArgValIleProArgMetGlyArgGlnLeuIle 1155
3480 ..... 3480
1156 AsnAspTyrGlnGlnIleProLeuGlnGlnLeuGlnAspGlnAlaGln 1172
3480 ..... 3480
1172 uGlnArgArgGlnAlaThrGlnPheHisSerLysSerArgAsnArgArg 1189
3480 ..... 3480
1189 laIleSerSerGlnProSerSerAspGlnAspAlaSerGlnSerValSer 1205
3480 ..... 3480
1206 ThrSerAspLysHisProGlnAspAsnThrGlnLeuHisGlnLysValGln 1222
3481 .....CTGCAACCCCAACCGCAGCCCAAC 3505
1222 uThrAlaGlnLeuGlnProArgAlaAlaGlnProArgThrGlnAlaAla 1239
3506 CGCAGCGCGACCTGATCAGCGCTTATGCAATAGCGTTGAGGAAATTT 3555
1239 laGlnAlaAspAlaValSerThrAsnThrAsnSerAlaLeuSerAla 1255
3556 TCCGACGCGTCAACAGCGT.....TTCGCGTACA 3587
1256 MetAlaSerThrGlnSerIleLeuLeuAspThrGlyAlaTyrLeuThrArg 1272
3588 GAGCAATTAAGACCGCGTATTTCGCAAGACCGCGCAACGCGCTTGA 3637
1272 gHisIleAlaGlnLysSerArgAlaAspAlaGlnLysAsnSerValTPrM 1289
3638 CAACCGCATCCGGGACACCAACACTACCTTCGCAAGATTTCCGCGC 3687
1289 eSerAsnThrGlyTyrGlyArgAspTyrAlaSerIleGlnTyrArgArg 1305
3688 TACCGCAACAAACCGACCTGCGCAATCGGATGATGAGAAACCTCGG 3737
1306 PheSerSerLysArgThrGlnThrGlnIleGlyIleAspSerLeuSe 1322
3738 CAGCGGG...CGCGTGGCATCCTGTTTCGCAACCGGACCGAAACA 3784
1322 rGlnAsnMetGlnIleGlyGlyValLeuThrTyrSerAspSerGlnHis 1339
3785 CTTTCACGACGCGATCGCACTCGGACGCGCTTCGCCAGCGCGCTT 3834
1339 hrPheAspGlnAla...GlyGlyLysAsnThrPheValGlnAlaAsnLeu 1354
3835 TTCGGCAATACGCGCATC...GACAGTTCTCATCGCATCGACGCGG 3881
1355 TyrGlyLysTyrTyrLeuAsnAspAlaTyrValAlaGlyAspIleGln 1371
3882 CGCGGCTTTAGCAGCGGACGCTTCAGACGCGCATCGGCAAAATCC 3931
1371 yAlaGlySerLeuArgSerArgLeuGlnThrGlnGlnLysAlaAsnPhe 1388
3932 GCGCGCGCTGCTGCATTCAGGATTCAGGACGACGATACCGCGCTTC 3981
1388 snArgThrSerIleGlnThrGlyLeuThrLeuGlnLysAsnThrLeuLysIle 1404
3982 GCGGATTCGATCGACCGCATCGGCGCAACCGCGCTATTGCTCA 4031
1405 AsnGlnPheGlnIleValProSerAlaGlyIleArgTyrSerArgLeuSe 1421
4032 AAAGCGATTACCGCTACGAAACGTCATATCGCACCGCGCGCTTG 4081
1421 rSerAlaAspTyrLysLeuGlyAspAspSerValLysSerMetAl 1438
4082 CATTCAACCGCTACCGCGCGGATTAAGCAGATATTCATTCAACCG 4131
1438 laValLysThrLeuThrAlaGlyLeuAspPheAlaTyrArgPheLys... 1453
4132 GCAGCAACATTCATTCATCACGCGCTTATTGACCTGCTCATACGATGC 4181
1454 ValGlyAsnLeuThrValLysProLeuLeuSerAlaIleAlaTyrPhe...Al 1469
4182 CGCTTGGCGCAAGTCCGACACGCGTCATATACCGCGCTATTGCTCAGG 4231
1469 aaSnTyrGlyLysGlyGlyValAsnValGlyGlyLysSerPheAlaTyrL 1486
4232 ATTTCGCAAAACCGCAGCGGATGGGCGGTAAACCGCAATCAAA 4281
1486 yAlaAspAsnGlnGlnGlnTyrSerAlaGlyValAlaLeuLeuTyrArg 1502
4282 GATTTCACGCTTCCTCCACGCTGCGCGCGCAAGGCGCGCATGCA 4331
1503 AsnValThrLeuAsnValAsnGlySerIleThrLysGlyLysGlnLeuGln 1519
4332 AGCGCAACACGCGCGGCGATCAATTAGGCTACGCGCTG 4371
1519 uLysGlnLysSerGlyGlnIleLysIleGlnIleArgPhe 1532
seq_name: SwissProt_40:IGA0_HAEIN
seq_documentation_block:
ID IGA0_HAEIN STANDARD: PRT: 1694 AA.

```


817 AsnAlaThrAsnValSerGlyAsnValAsnLeuSer 828
2304 AAsrTCTACAGGGCTTGGCACATCTACAGGCAATCTTAGTCAAAATGCGC 2353
828 828
2354 ATACAGCTTATACAGTACGCCAACAGCCCAAAAGCGCACTTACG 2403
828 828
2404 CTCTGGGCAATGCCCAAGCAACATTTATCAGCCACATTAAGCGCAA 2453
829 GlyAsnAlaAsnPhenValLeuGlyLysAlaAsnLeuPheGlyTh 843
2454 CACATCGGCTGGGCAATGCTTCAATTAATCTAAGCGACGACCGCGTAC 2503
843 rLleSerGlyThrGlyAsnSerGlnValAlaArgLeuThrGlu..... 856
2504 AAAAGCGAGCTGTGACGCTTCCGGCAACGCTAAGCAAGTAAAGCAT 2553
856 856
2554 TCCGCACTCAACGTAATGTCTCCTAGCGATAAGGAGTATTCATTT 2603
856 856
2604 TGAAGACAGCCGCTTACCGGCAATACAGCGCGGCAAGATACGCGAT 2653
856 856
2654 TACACTTAAAGACAGACGTAATGACGCTGCGCTACGACGCAATTAAGC 2703
857 AsnSerHisThrPheLeuThrGlyAspSerAsnValAsn 869
2704 AATTAAACCTTTGACACAGCCACCATTTACACTCAATTCGCGCTAACGCCA 2753
870 GlnLeuAsnLeuAspLysGlyHisIleHisLeuAsnAla..... GlnAs 884
2754 CGATGGGCGAGGGGCGCAACCGGCAAGTGCAGACAGATGGCGCGCGCCGC 2803
884 nAspAlaAsnLysValThrThr..... 891
2804 GTTCGGCGCTTCGGCGCGCTCCATTATTCGTTACACCGCAACTTCG 2853
891 891
2854 GTAGAATCCCGTTTTCACACGCTGACGTAACGGCAATTTAAGGTCA 2903
892 TyrAsnThrLeuThrValAsn...SerLeuSerGlyAs 903
2904 GGGACATTCGCGTTTATGTGGAACCTTCGCTACCGCAAGCGACGCAAT 2953
903 nGlySerPheTyrTyrLeuThrAspLeuSerAsnLysGlnGlyAspLysV 920
2954 TGAAGCTGGCGGAAAGTTCCGAAGCATTTACACTTCGCGCTCAACATV 3003
920 aValValThrLysSerAlaThrGlyAsnPhenThrLeuGlnValAlaAsp 936
3004 ACCGGCAAGCACTGCAAGCTCGAACAATTGACGCTAGTGGAAAGGAAA 3053
937 LysThrGlyGlnProThrLys...AsnGluLeuThrLeuPheAspAla... 951
3054 AGACAACAACCGCTGTCCGAAAACCTTAATTTCAACCTGCAAAAAGAAC 3103
952 SerAsnAlaThrAlaGlnAsnLeuAsnValSerLeuValGlyAsnT 967
3104 ACCTGCATGCGCGCGCTGACCACTATCCGCAAGACGCGCGAG 3153
967 hrValAspLeuGlyAlaThrLysTyrLysLeuAlaArgAsnValAsnGlyArg 983
3154 TTCCGCTGCAATTAATCG..... 3171
984 TyrAspLeuTyrAsnProGlnValGlyLysArgAsnGlnThrValAspTh 1000

3171 3171
1000 rThrAsnIleThrThrProAsnAsnIleGlnAlaAspValProSerValP 1017
3171 3171
1017 roSerAsnAsnGlnGluIleAlaArgValGluThrProValProPro 1033
3171 3171
1034 AlaProAlaThrProSerGluThrThrGluThrValAlaGluAsnSerLy 1050
3171 3171
1050 sGlnGluSerLysThrValGluLysAsnGlnAspAlaThrGluThrT 1067
3171 3171
1067 hrAlaGlnAsnGlyGluValAlaGluGluAlaLysProSerValLysAla 1083
3171 3171
1084 AsnThrGlnThrAsnGlnValAlaGlnSerGlySerGluThrGluGluTh 1100
3172 GTCAAAGAA..... 3180
1100 rGlnThrThrGluIleLysGluThrAlaLysValGluLysGluLysAla 1117
3181 CAAGAG 3186
1117 IalysValGluLysAspGluIleGlnGluAlaProGlnMetAlaSerGlu 1133
3187 CTTTCCGACAACCTCGGCAAG..... 3207
1134 ThrSerProLysGlnAlaLysProAlaProLysGluValSerThrAspTh 1150
3207 3207
1150 rLysValGluGluThrGlnValGlnAlaGlnProGlnThrGlnSerThrT 1167
3208 GCAGAAGCC 3216
1167 hrValAlaIleAlaGluAlaThrSerProAsnSerLysProAlaGluGlu 1183
3217 AAAAAGACGCGGAAAAAGACAAACGCGCAAGCCTTGACGCGGTATGC 3266
1184 ThrGlnProSerGluLysThrAsnAlaGluProValThrProValAlaSe 1200
3267 G.....G 3268
1200 rLysAsnGlnThrGluAsnThrThrAspGlnProThrGluArgGluLysT 1217
3269 CCGGGCGCGATCGCGTGAAGAAAGACAGACGCTGCGGAACGCGCGCG 3318
1217 hrAlaLysValGluThrGluLysThrGlnGluProProGlnValAlaSer 1233
3319 CAGCGAGGC..... 3327
1234 GlnAlaSerProLysGlnGluGlnSerGluThrValGlnProGlnAlaVala 1250
3328 GGGGAAATGTC...GGCATTATGCAAGCGGAGAGAGAAA 3367
1250 IleuGlnSerGluAsnValProThrValAsnAsnAlaGluGluValGlnA 1267
3368 AAGGAGTGCAGCGCATTAAGAC..... 3390
1267 IagInLeuGlnThrGlnThrSerAlaThrValSerThrLysGlnProAla 1283
3391 ACCGCTTGGCGAAACAGCG 3410
1284 ProGluAsnSerIleAsnThrGlySerAlaThrAlaThrGluThrAl 1300

```

3411 CGAAGCGAACCAGCCGGCTAC..... 3435
1111 : : : : :
1300 aglutysserapryspruginhrthralaaserthgluasp 1317
3435 ..... 3435
1317 laseglnhislysalasnthvalaaspasnservalalaasnsp 1333
3436 ..... ACCGCTTCCCGCGCGCGCGCGCGCGGATTTGCCGA 3479
: : : : :
1334 sergluserasprprolysserarglyg... Argargserllesergl 1349
3480 ACTGCAA..... 3486
1111 : : : : :
1349 nproglngluthrseralaglulthrthralaaserthaspolut 1366
3487 ..... CCGAACCGGACCGCCACCGCGAG 3510
: : : : :
1366 hrthrlleaaspasnserylsergserlyspasnarargserarg 1382
3510 ..... 3510
1383 ArgSerValArgsergluprothrValThrAsnGlySerAspArgserth 1399
3511 ..... CCGGACCTGATCAGCGCTTATGCCAATAGCGGTTGAGTG 3550
1111111111 : : : : :
1399 rValAlaLeuArgAspLeuthrSerThrAsnThrAsnAlaValIleSera 1416
3551 AATTTTCCGCC..... ACCTCAACAGCGCTTTTCCGCCGA 3585
: : : : :
1416 sprlmetrAlaLysAlaGlnPheValAlaLeuAsnAlaLylLysAlaVal 1432
3586 CAGAGCAATTAAGCCGCTA... TTTGCCAAGACCGCGCGACGCCCT 3632
: : : : :
1433 SerGlnHisIleSerGlnLeuGlnMetAsnGlnGlyGlnThrAsnVA 1449
3633 TTGGACACCGGGATCCGGACACCAACACTACCGCTTCCGCAAGATTTC 3682
1111111111 : : : : :
1449 ltrpValSerAsnThrSerMetAsnGlnLysThrSerSerGlnTYA 1466
3683 GCGCCTACCGCCACCAACAGCACTGCGCAATCGGTATGCAAGAAAC 3732
: : : : :
1466 rGArgPheSerSerlysserThrGlnThrGlnLeuGlyTTPAspGlnThr 1482
3733 CTCGGCAGCGGG... CGGTCGGCATCTGTTTGCACACCGGACCGA 3779
: : : : :
1483 lIleSerAsnAsnValGlnLeuGlyValPheThrTYValArgAsnSe 1499
3780 AAACACCTTGACGACGCGATCGGACACTGCGGACGCTTCCCGACGGCG 3829
: : : : :
1499 rAsnAsnPheAspLysAlaSerSerLysAsnThr... LeuAlaGlnValA 1515
3830 CCGTTTGGGGCAATACGGCATCGAC... AGTTTCTACATCGGCATCAGC 3876
: : : : :
1515 snPheTYSerlystyTYrAlaAspAsnHisTPTTYLeuGlyLeasp 1531
3877 GCGGCGCGGGGTTTACGACGGGACGCTTTCACAGCGCATCGAGGCAA 3926
1111 : : : : :
1532 LeuGlyTYrGlyLysPheGlnSerAsnLeuLysThrAsnHisAsnAlaLy 1548
3927 AATCGCGCGCGCGGCTGATACGGCATTCAGGACGACGATACCGCGCGG 3976
: : : : :
1548 sPheAlaArgHisThrAlaGlnPheGlyLeuThrAlaGlyLysAlaPheA 1555
3977 GTTTCGGCGGATTCGCATCGAACCGACATCGGCGACCGCTATTTC 4026
: : : : :
1565 snLeuGlyAsnPheGlyIleThrProIleValGlyValAlaTYrTYr 1581
4027 GTCCAAAAGCGGATACCGCTACGAAAAGCTCATATTCGACACCGCGG 4076
: : : : :
1582 LeuSerAsnAlaAsnPheAlaLeuAlaLysAspArgTlleLysValAsnPr 1598
4077 CATTGATTCACCGCTACCGCGCGGCGCATTAAGCAGATTAATTCATTCA 4126

```

```

1598 orIleSerValLysThrAlaPheAlaGlnValAspLeuSerTYrThrTYrH 1615
4127 AACCGCGCACACATTTCCATCAGCGCTTATTAGCTGTCCAT... 4173
: : : : :
1615 lLeuGlyGln... PheSerValThrProIleLeuSerAlaArgTYrAsp 1630
4174 ACCGATGCCGCTTTCGGGCAAAATCCGAAACAGCGCTTATACCGCGTAT 4223
: : : : :
1631 ThrAsnGlnGlySerGlyLysIle... AsnValAsnGlnTYrAspPh 1645
4224 GGCTCAGGATTTCCGCAAAACCGGACATGCGGATGGCGCTAAACGCCG 4273
: : : : :
1645 eAlaTYrAsnValGlnAsnGlnGlnTYrAsnAlaGlyLeuLysLeuL 1662
4274 AATCAAGGTTTCAAGCTTCCTCCAGCTGCGCGCGCGCAAGGCCCG 4323
: : : : :
1662 ySTYrHisAsnValLysLeuSerLeuIleGlyGlyLeuThrLysAlaLys 1678
4324 CAACTGGAAGCGCAACACAGCGCGGCAATTAAGGCTAC 4365
1111111111 : : : : :
1679 GlnAlaGlnLysGlnLysThrAlaGlnLeuLysLeuSerPhe 1692

```

seq_name: SwissProt_40:IGA2_HAEIN

seq_documentation_block:

ID IGA2_HAEIN STANDARD; PRT; 1702 AA.

AC P45384;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).

GN IGA.

OS Haemophilus influenzae;

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OX Haemophilus;

NCBI_TaxID=727;

RP [1]

RF SEQUENCE FROM N.A.

RC STRAIN=HK715 / SEROTYPE B;

MDLINE=92234949; PubMed=1373717;

RA Foulson K., Reinholdt J., Killian M.;

RT "A comparative genetic study of serologically distinct Haemophilus

influenzae type 1 immunoglobulin A1 proteases.";

RL J. Bacteriol. 174:2913-2921(1992).

CC - FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A

PRODUCING INTACT FC AND FAB FRAGMENTS.

CC - CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at

certain Pro-I-Xaa bonds in the hinge region. No small molecule

CC substrates are known.

CC - SUBCELLULAR LOCATION: Secreted.

CC - DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC

SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE

OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE

DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY

SIMILARITY).

CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC EMBL; M87489; A24966.1; -

DR MEROPS; S06.001; -

DR InterPro; IPR000710; IGA_S6.

DR Pfam; PF02395; IGA1; 1

DR PRINTS; PR00921; IGA5ERPTASE.

Hydrolase; Serine protease; Transmembrane; Zymogen; Repeat; Signal.

KM SIGNAL 1 25 POTENTIAL.


```

492  sAlaPheSerGlnValGlyIleValSerGlyArgSerThrValIleLeu 509
1547  ATGCCGATTAATCAGTTCAACCCGACAACTCATTTCCGCTTCCGGC 1596
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
509  snmSapRpySgInValAspProAsnSerIleThrPheGlyPheArgGly 525
1597  GGAAGTTTGGATTTAAACGGGCACTTCCTTCCTCCACCGATATCAAAA 1646
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
526  GLyArgLeuSapAlaAsnGlyAsnAsnLeuThrPheGlnHisIleArgS 542
1647  TACCGGATGAAGGGCGCATGATGTTCACACCAATCAAGACAAAGATCCA 1696
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
542  nIleAspRpyGlyAlaArgLeuValAsnHisAsnThrSerLysThrSer 559
1697  CCGTTACCATTCAGGC..... 1713
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
559  hValThrIleThrGlyGlySerLeuIleThrAspProAsnThrIleThr 575
1713  ..... 1713
576  ProTyrAsnIleAspAlaProAspGluAspAsnProTyrAlaPheArg 592
1713  ..... 1713
592  gIleLysAspGlyGlyGlnLeuTyrLeuAsnLeuGluAsnTyrThrTyr 609
1713  ..... 1713
609  yValLeuArgLysGlyAlaSerThrArgSerGluLeuProLysAsnSer 625
1714  ..... 1727
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
626  GLyGluSerAsnGluAsnTrpLeuTyrMetGlyLysThrSerAspGluAl 642
1728  TACAACCGCGCAATTAACACAGCTGGATGCAAAAACAAATGGCTACA 1777
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
642  aLysArgAsnValMetAsnHisIleAsnSngluArgMetAsnGlyPheA 659
1778  ACGGTGGTTGGCGAGAAAGATACGACCAACAAACGACGGCGGCTCAAC 1827
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
659  sngIlyThrPheGlyGlnGluGlu...GlyLysAsnSnglyAsnLeuAsn 674
1828  CTGTGTTACCGACCGCGGCAACACCGCACTTCGCTTCCGGCGG 1877
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
675  ValThrPheLysGlyLysSerGlnGluAsnArgPheLeuLeuThrGlyG 691
1878  AACAAATTTAAACGGCAATCATCACGCAAAACAAACGCAAACTGTTTCA 1927
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
691  yThrAsnLeuAsnGlyAspLeuLysValGlnLysGlyThrLeuPheLeu 708
1928  GCGGACAGACCAACCGCACCGCTACAA..... 1956
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
708  eGlyArgProThrProHisAlaArgAspIleAlaGlyIleSerSerThr 724
1957  CATTTAAACGACCATTTGGTGGCAAAAGAGGGCAATTCCTCGCGGGAAT 2006
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
725  LysLysAspGlnHisPheAlaGlu.....AsnAsnGluVal 736
2007  CGTGGGAGACAGACTGGATCAACCGCACATTTAAAGCGGAAACTTCC 2056
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
736  lValValGluAspArgPheIleAsnArgAsnPheLysAlaThrAsnIleA 753
2057  AAATTAAGGGGAGAGCGCGGTGTTCC...CCGATGTTCGCAAAAGTG 2103
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
753  sNValThrAsnAsnAlaThrLeuTyrSerGlyArgAsnValAlaAsnIle 769
2104  AAAGCGATTTGGCATTTGAGCAATCAGCCCAACGATTTTGGTGTGCG 2153
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
770  ThrSerAsnIleThrAlaSerAspAlaLysValHisIleGlyTyr... 785
2154  ACGCGATCAAGGCCACACATCTTACACGTTGCGACTGACGGGTCTGA 2203
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
786  ...LysAlaGlyAspThrValLysValArgSerAspTyrThrGlyTyr 801
492  sAlaPheSerGlnValGlyIleValSerGlyArgSerThrValIleLeu 509
2204  CAAATTTGTGTCGAAAAAACCATTAACCGACGATTAAGTATGCTTCAT 2253
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
801  aThrCysThrThrAspLysLeuSer...AspLysAlaLeuAsnSerPhe 816
2254  ACTAAGACCGACATCAGCGGCAATTCGATCTTGCCATCAAGCTCATTT 2303
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
817  AsnAlaThrAsnValSerGlyAsnValAsnLeuSer..... 828
2304  AAATCTCACAGGGCTTGCACACATCAACGGCAATCTTAGTCAAATGGCG 2353
828  ..... 828
2354  ATACAGTTTATACAGTCAGCACCAACGCCAACCAAAACGGCACTTAGC 2403
828  ..... 828
2404  CTCGTGGGCATTCGCCAACCAACATTTATACAGCCACATTTAAACGCA 2453
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
829  ...GlyAsnAlaAsnPheValLeuGlyLysAlaAsnLeuPheGlyTh 843
2454  CACATCGGCTTCGGGCAATGCTTCATTTAATCTAAGCGACACCGCGTAC 2503
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
843  rIleSerGlyThrGlyAsnSerGlnValArgLeuThrGlu..... 856
2504  AAAACGGCATCTGACGCTTTCGGCAACGCTAAGCAAAACGTAAAGCAT 2553
856  ..... 856
2554  TCCGACTCAACGGTAATGTCTCCCTAGCCGATAGGACATTTCCATTT 2603
856  ..... 856
2604  TGAAGACAGCGCTTACCGGACAAATCAGCGGCGCAAGATACGGCAT 2653
856  ..... 856
2654  TACACTTAAAGACAGAGGAATGACGCTGCCCTCAGACGCAATTAAGC 2703
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
857  ...AsnSerHisIleThrPheIleuThrGlyAspSerAsnValAsn 869
2704  AATTTAAACCTTGACACAGCGCACTTACACTCAATTCGCGCCATAGCCA 2753
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
870  GlnLeuAsnLeuAspLysGlyHisIleHisIleAsnAla.....Glns 884
2754  CGATGCGCGAGGGGCAACCGGCACTGCGACAGATGCGCGCGCGGCC 2803
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
884  nAspAlaAsnLysValThrThr..... 891
2804  GTTCGCGCGCTTCGCGCGCTTCCTATATCCGTACCGCGCAACTCG 2853
891  ..... 891
2854  GTAGAAATCCGTTTCAACAGCGTACGAGTAACGCAAAATGAAGCGTA 2903
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
892  ...TyrAsnThrIleuThrValAsn...SerLeuSerGlyAs 903
2904  GGAACATTCGCGTTATATGCGAAGCTTCGCTTACCGGACGAGCAAT 2953
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
903  nGlySerPheTyrTyrLeuThrAspLeuSerAsnLysGlnGlyAspLys 920
2954  TGAAGCTGGCGGAAGTTCGGAAGCACTTACCTTGGCGGCAACAT 3003
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
920  aValValThrLysSerAlaThrGlyAsnPheThrLeuGlnValAlaAsp 936
3004  ACGGCAACGACCTCAAGCCTCGACAAATTCAGCGGTAGTGAAGGAAA 3053
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
937  LysThrGlyGluProThrLys...AsnGluLeuThrLeuPheAspAla.. 951
3054  AGACAAACAACCGCTGTCGAAACCTTAATTCACCCCTGCAAAAGCAAC 3103
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
952  ...SerAsnAlaThrArgAsnAsnLeuAsnValSerLeuValGlyAsnT 967

```

```

3104 ACSTGATGCGCGCGGTACCACTACCGCAAGAGCGGAG 3153
      ||||| ||||| ||||| ||||| ||||| |||||
967 hrValAspLeuGluAlaIrrPylusLeuArgAsnValAsnGlyArg 983
      : : : : : : : : : : : : : : : : : :
3154 TTCCGCGTGCATATCCG..... 3171
      : : : : : : : : : : : : : : : : : :
984 TyrAspLeuTyrAsnProGluValGluLysArgAsnGlnThrValAspTh 1000
3171 ..... 3171
1000 rThrAsnIleThrThrProAsnAsnIleGlnAlaAspValProSerValP 1017
3171 ..... 3171
1017 roSerAsnAsnGluGluIleAlaIrrValGluThrProValProProPro 1033
3171 ..... 3171
1034 AlaProAlaThrProSerGluThrThrGluThrValAlaGluAsnSerLy 1050
3171 ..... 3171
1050 sGlnGluSerLyThrValGluLysAsnGluInAspAlaThrGluThrT 1067
3171 ..... 3171
1067 hrAlaGlnAsnGlyGluValAlaGluGluAlaLysProSerValLysAla 1083
3171 ..... 3171
1084 AsnThrGlnThrAsnGluValAlaGlnSerGlySerGluThrGluGluTh 1100
3172 .....GTCAAGACAAAGAGCTTCCGACAAACTCGCAGC 3208
      : : : : : : : : : : : : : : : : : :
1100 rGlnThrThrGluIleLysGluThrAlaLysValGluLysGluGluLysA 1117
3209 CAGAAGCAAAAAACAG..... 3225
      || : : : : : : : : : : : : : : : : : :
1117 IalysValGluLysGluLysAlaLysValGluLysAspGluIleGln 1133
3225 ..... 3225
1134 GluAlaProGlnMetAlaSerGluThrSerProLysGlnAlaLysProAl 1150
3225 ..... 3225
1150 aProLysGluValSerThrAspThrLysValGluGluThrGlnValGlnA 1167
3225 ..... 3225
1167 IagInProGlnThrGlnSerThrThrValAlaAlaIleGluAlaThrSer 1183
3226 .....GCGAAGAAAGACAAAGC 3242
      : : : : : : : : : : : : : : : : : :
1184 ProAsnSerLyProAlaGluGluThrGlnProSerGluLysThrAsnAl 1200
3243 GCAAGCCTTGACGCGTGAATGCG..... 3267
      : : : : : : : : : : : : : : : : : :
1200 aGluProValThrProValIleSerLysAsnGlnThrGluAsnThrThra 1217
3268 .....GCGGGCGCGATCGCGTGAAGAGCA 3294
      : : : : : : : : : : : : : : : : : :
1217 spGlnProThrGluArgGluLysThrAlaLysValGluThrGluLysThr 1233
3295 GAAAGCGTTGCGAAGCCGCGCGAGCAGC..... 3327
      : : : : : : : : : : : : : : : : : :
1234 GlnGluProProGlnValAlaSerGlnAlaSerProLysGlnGluInSe 1250
3328 .....GCGGAAATGTC...GSCA 3343
      : : : : : : : : : : : : : : : : : :
1250 rGluThrValGlnProGlnAlaValLeuGluSerGluAsnValProThrV 1267
3344 TTATCGAGCGAGAGAGAAAAAAGCGGTGCGAGCGGATTAAGAC... 3390

```

```

      : : : : : : : : : : : : : : : : : :
1267 aIAsnAlaIleGluGluValGlnAlaGlnLeuGlnThrGlnThrSerAla 1283
3390 ..... 3390
1284 ThrValSerThrLysGlnProAlaProGluAsnSerIleAsnThrGlySe 1300
3391 .....ACCGCTTGCGAAGACGCGAAGCGGAAACCGCGCGGTACC. 3435
      ||||| : : : : : : : : : : : : : : : : : :
1300 rAlaThrAlaIleThrGluThrAlaGluLysSerAspLysProGlnThrG 1317
3435 ..... 3435
1317 IuThrAlaIleSerThrGluAspAlaSerGlnHisLysAlaAsnThrVal 1333
3436 .....ACCGCTTCCCGCGCGCG 3455
      : : : : : : : : : : : : : : : : : :
1334 AlaAspAsnSerValAlaAsnAsnSerGluSerGluProLysSerAr 1350
3456 CCGCGCGCGCGGATTTGCGCAACTGCA..... 3486
      ||||| : : : : : : : : : : : : : : : : : :
1350 garG...ArgArgSerIleSerGlnProGlnGluThrSerAlaGluGluT 1366
3486 ..... 3486
1366 hrThrAlaIleSerThrAspGluThrThrIleAlaAspAsnSerLysArg 1382
3487 CCCCACCGCGAGCCCGACCGCAG..... 3510
      : : : : : : : : : : : : : : : : : :
1383 SerLysProAsnAlaArgArgSerArArgSerValArgSerGluProThrVa 1399
3511 .....CGCGACCTGATCAGCC 3526
      ||||| : : : : : : : : : : : : : : : : : :
1399 IThrAsnGlySerAspArgSerThrValAlaLeuArgAspLeuThrSerT 1416
3527 GTATGCAATAGCGGTGAGTGAATTTCCGCC..... 3561
      : : : : : : : : : : : : : : : : : :
1416 hrAsnThrAsnAlaValIleSerAspAlaMetAlaLysAlaGlnPheVal 1432
3562 ACAGCTACAGCGGTTTCCCGTACGAGCAATTAAGCCGCTA...TT 3608
      : : : : : : : : : : : : : : : : : :
1433 AlaLeuAsnValGluLysAlaValSerGlnHisIleSerGlnLeuGlnMe 1449
3609 TGCCGAAGACCGCGGACCGCGTGGACAGCGGACATCCGGACACCA 3658
      : : : : : : : : : : : : : : : : : :
1449 IAsnAsnGluGlyGlnTyrAsnValTyrPalaSerAsnThrSerMetAsnG 1466
3659 AACACTACCGTTCGCAAGATTTCCGCGCTACCGCAACAAACCGACTG 3708
      : : : : : : : : : : : : : : : : : :
1466 IuAsnTyrSerSerSerGlnTyrArgArgPheSerSerLysSerThrGln 1482
3709 CGCCAAATCGTATCGAAGAAACCTCGGACGCGG...CGCGTCGCGAT 3755
      ||||| : : : : : : : : : : : : : : : : : :
1483 ThrGlnLeuGlyTrrAspGlnThrIleSerAspAsnValGlnLeuGly 1499
3756 CCGTTTTCGACACCGCGGCAAAACACTTCGACGACGACGATCGGCA 3805
      : : : : : : : : : : : : : : : : : :
1499 yValIrrPheThrTyrAlaArgAsnSerAsnPrleAspLysAlaSerSerL 1516
3806 ACTGCGACGCGTTCGCGAGCGCGCGTTCGCGCAATACGCGATCGAC 3855
      : : : : : : : : : : : : : : : : : :
1516 yAsnThr...LeuAlaGlnValAlaAsnPheTyrSerLysTyrAlaAsp 1531
3856 ...AGTTTACATCGGATCGACGCGCGCGGCTTTAGCAGCGGAG 3902
      : : : : : : : : : : : : : : : : : :
1532 AsnHisIrrPyrLeuGlyIleAspLeuGlyTyrGlyLysPheLeuSerAs 1548
3903 CCGTTGACGCGATCGAGCAAAATCGCGCGCGGTGCTCATTAAG 3952
      : : : : : : : : : : : : : : : : : :
1548 pleuLysThrAsnHisAsnAlaLysPheAlaArgHisThrAlaGlnPheG 1565
3953 GCATTGACGCGATACCGCGCGGTTCGCGGATTCGCGATCGAAGCG 4002
      || : : : : : : : : : : : : : : : : : :

```

```

1565 lyleuthralaglylsalaphnealsnleuglyasnphgilylerthrpro 1581
4003 CACATCGCGCCGACGCGCTATTGTCACAAAGCGGATTACCGCTTCGCA 4052
1582 lilevalglyvalargtyrserlyrleuserasnlaasnphalaleual 1598
4053 AAGCTGAATATCGCCACCCCGCGCTTGATTCACCGCTACCGCGG 4102
1598 alysaspargilelysalasnproilservallysthralaphnealag 1615
4103 GCATTATGGAGATTATTCATTCAACCGCGCACACATTTCATCAGC 4152
1615 lvalaspleuserlyrthrlysthisleuglyglu...pheservalthr 1630
4153 CCTATTATGGAGCTGCTCTAT...ACCGATGCGCGCTGCGCAAGCGC 4199
1631 proileuseralalargtyraspthrphasnnglyserglystle... 1646
4200 AACACGCGCTAATACCGCGCTATTGCTCAGATTCGCGCAAAACCGCA 4249
1647 ...AsnValasnnglnlyrasphthalatyrasnvalgluasnnglng 1662
4250 GTCCGGAATGGGCGGTAAACGCGCAATCAAGTTTCACGCTGCTC 4299
1662 lntyrasnlaaglyleulysleulysrtyrhisasnvallylseuSerleu 1678
4300 CACGCTCGCGCGCCGCAAGCGCGCACTGGAAGCGCAACACACGCGG 4349
1679 lileglyllyleuthrlysalatylsglnalaglylsglysthralag 1695
4350 CATCAATTAAGGCTAC 4365
1695 uleulysleuserphe 1700

seq_name: SwissProt_40:IGA4_HAEIN

seq_documentation_block:
ID IGA4_HAEIN STANDARD: PRT: 1849 AA.
AC P45386;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NHTI HK61;
RX MEDLINE-92234949; Pubmed-1373717;
RA Poulsen K., Reinholdt J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
RT influenzae type 1 immunoglobulin A1 proteases.";
RL J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTERACT EC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at
CC certain pro-1-xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE. AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M87491; AAA24968.1; -
DR MEROPS; S06.001; -
DR InterPro; IPR000710; IGA-S6.
DR Pfam; PF02395; IGA1; 1.
DR PRINTS; PR00921; IGASERPTASE.
KW Hydrolase; Serine protease; Transmembrane; zymogen; signal.
FT SIGNAL 1 25
FT CHAIN 26 1021 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1022 1849 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 299 299 PROBABLE.
FT SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150AFA CRC64;

alignment_scores:
Quality: 1488.50 Length: 1917
Ratio: 1.493 Gaps: 59
Percent Similarity: 52.008 Percent Identity: 24.935

alignment_block:
US-09-303-518D-649 x IGA4_HAEIN ..

Align seg 1/1 to: IGA4_HAEIN from: 1 to: 1849

64 CGCTTCGCGCTGCTTACTTACCATATGCTGCTGCTGCGCATTCCTCC 113
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
5 LysPheLysLeuAsnPhelaleuthrValAlatyrAlatleuthrpr 21
114 CCAAGCTGGGCGGACACTTATTTCGCGATCACTACCATATCATC 163
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
21 OtyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38
164 GCGACTTTCGCGAAATTAAGCAAGTTTCGATCGGCGCAAGATATT 213
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
38 rgAspPheAlaGluAsnLysGlyLysPheSerValGlyAlatThrAsnVal 54
214 GAGGTTTCACAAAGGAGGAGTGTGCGCAATCATGACAAAGC 263
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
55 GluValArgAspLysLysAsnGlnSerLeuGlySerAlaLeuProAsnG 71
264 C...CGATGATGATTTTCTGTGTGTG...CGTAAGCGCGTGGCG 307
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
71 YIleProMetIleAspPheSerValValAspValAspLysArgIleAlat 88
308 CATTTGGTGGCGATCAATATATTGTGACGCTGCACAT...AAGCGCGC 354
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
88 hrLeuValAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104
355 TATAACAAGTGTGATTTTGGT...GCGGAGAGAGAAA 389
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnGlnGly 121
390 TCCGATCAACATCGTTTACT...TATAAATG 421
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
121 nAlaLysSerHisArgAspValSerSerGluGluAsnArgTyrTyrThy 138
422 TGAACGCAATTAATT...AAA 441
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
138 alGluLysAsnAsnPheProThrGluAsnValThrSerPheThrLys 154
442 GCAAGGCAATAAGGCATCTTATGGCGGATTAATCATATGCGGGTT 491
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
155 GluGluGlnAspAlaGlnLysArgArgGluAspTyrTyrProArgIle 171
492 GCATTAATTTGTCACAGATCAGACACCTGTTAAAGACAGTTATATG 541
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
171 uAspLysPheValThrGluValAlaProIleGluAlaSerThrAlaAsn 188
542 ATGGCGG...AATATATCATCAAAATTAATTAACCTGACCGGTGCT 588
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```
188 snbsnlysglylutyraasnaserasplystyrproalaphavalarg 204
589 ATGGGGCAGCAGCAATAT..... 609
205 leuglyserglyserglnpheiletyrlylysglyserargtyrclnle 221
610 .TGCGCATCTGATGAGATGAGCCCAATACCCGGAAGTTCATATCAT. 657
221 uileleuthrclnlylsasplysglnglnleuauagasntrpsrv 238
658 ..... ATGCAGTCGCTATTCCTGGCTGCT 684
238 algllylaspaslnleuglnleuvalgllysnhlatythrlyrlygylle 254
685 GGTGCAATACCTTTGCACAAATGATGATGAGGTGGCAGCTCAACTT 734
255 Alaglythrprotyrlylsvalasnhsisgluasnbnllyleuileglyph 271
735 AGGTAGGAAAAATTAACATAGC...CCATAGGTTTTTTTACCAACAG 781
271 eglysnserlysglnlunhsiserasprolysglylleuuserclna 288
782 GA.....GGTCATTTGGCGCAGACAGTGGCTCACCAGTGT 816
288 sprproleuthrasnlyralavalleuglyaspserylserproleuphe 304
817 ATGTATGATGCCAAAGCAAAAGCTGTTATATGAGGTATTTGCAAC 866
305 Valtyrasparaglyllylstrpleupheleuglyserlyrasph 321
867 GCGCAACCCCTATATAGAAAAAGCAATGGCTCCAGCTGTTGTTAAAG 916
321 etrpalacllytyrasnlylserlytrpnlclntrpasnlelyrlysh 338
917 ATTGGTTCTATGATGAATCTTTGCTGAGATACCATTAGATATCTAC 966
338 lsglnphealaglulysllytyr.....Glnglntyrserralaglyser 352
967 CAACCACTCAAAATGGCAATCTTTTACGACGATATATATGGCAC 1016
353 leuthrlyserasnlyrlnlytyrthrtp.....Glnalath 365
1017 AGGAAAAATCATGCGCAACATGAACATTCCTGCTAATGATTA 1066
365 rgly.....S 367
1067 AAACGACGACCTT.....CATTTGTTAATGTTCTTTATCC 1104
367 erttserthrlethrlygllygluproleuservalaspleuthr 383
1105 GAGACAGCAGAGACCTGTTATCATGCTGAGGTGTCACACAGTTA 1154
384 aspglylylsasplyspro..... 389
1155 TCGACCCAGACTGATTAATGAGAAATATTTCTTTATGACGAGGAA 1204
390 .....Asnhsisglylserlethrleu.....lysglys 400
1205 AAGCGAATGCTACTTACGACGACATCATCAAGGTGGGAGGATTA 1254
400 ertglythrleuthrleuasnbnhsisileasplnglylaleglylyleu 416
1255 TATTTCCAGAGATTTTACGCTC...TCGCTGAAATTAACGAACCTTG 1301
417 phephegluglyasptyrclnvallylsglythrseraspsertthrthr 433
1302 GCAAGCGCGGCGTTCATATCATGAGACAGTACCGTTACTTGTGAAG 1351
433 plysglyalaglylaservalalaspolylystrvalaThrtrplysv 450
1352 TTAAGCGCTGCAAAAGCGCGCTGTCCAAATTCGCAAAAGCGCGCTG 1401
450 alhtisnprolystytraspargleuvalalyleglylysglythrleu 466

1402 CACGTTCAAGCCAAAGGGAACCAAGCTCTGATCAGCGTGGCGACGG 1451
467 valValgluglyllysglyllyasnlnclnlyleuenuysvalglaspgl 483
1452 TACACTCATTTTGGATCAGACGAGCAGACGATTAAGGCAAAAACAAGCTT 1501
483 ythrvalilleuulysglnlysalaspsalasnbnlylsvalglinalap 500
1502 TTATGGAATCGGCTTGTCACCGCGGAGGTACGCTCACTGAATGCC 1551
500 heserclnvallyllevalserglylaserthrleuvalleuasnasp 516
1552 GATATCATCTTCAACCCGACCAAACTATTTGCGCTTTCGCGGCGACG 1601
517 asplysglnvalasproasnserllytyrphelyphearglyglylar 533
1602 TTTGATTTTAAAGGCGCATTCGCTTTCGTCACCGCTATTCATAATACCG 1651
533 gleuaspleuasncllyasnserleuthrphasnhsisilearasnillea 550
1652 ATGAAGGCGCGATGATTCACACCAATCAAGACAAAGAAATCCACGCT 1701
550 spaspolyalalyvalalvalasnhsisnmetlthrasnthserasnille 566
1702 ACCATTACAGC.....ATAAAGATATTGCTACACCGG 1736
567 thrllethrclglyluserleuthrasnproasntrhlethrserly 583
1737 CAAT..... 1740
583 rsnlileglualaglnaspsasphlsproleuarglleargserllep 600
1740 ..... 1740
600 rotyrarglnleuthrphasnlnasnasnargserlytyrthrleu 616
1740 ..... 1740
617 lyslysglylaserthrargsergluleuproglinasnserglylgluse 633
1740 ..... 1740
633 rasngluasntrpleuthrmetelylarythrserasplalalyasga 650
1741 .....AACACAGCTTGATAGCAAAAAGAAATTCCTACACAGGCTGG 1785
650 snvalmetasnhsisileasnangluargmtasncllypheasnlytyr 666
1786 TTTGGCGAGAAAGATACGACCAAAAGC...AACGGCGGCTCAACCTGT 1832
667 pheglyglunclnlythrlysalatrlnasnglylylsleuasnvalth 683
1833 TTACCGCCCGCGCGAGAAAGCCGACCCCTGCTTTCGCGGAGAACAA 1882
683 rphenasnglyllyserasplasnargrphleuclnlygllythra 700
1883 ATTTAAACGCGCAACATCAGCAAAACAAAGGCAAACTGTTTTCACCGG 1932
700 snleuasnnglyllyaspleuasnvalglulysglythrleupheleusergly 716
1933 AGACCAACACCGCGACGCTCATCATCTTAAAGGACCATTTGGCGGAAAA 1982
717 Argprothrprohlsalatrpsalatrlnaleaglylleserthrlyslly 733
1983 AGAGGCG.....ATTCTCGCGGGAATCGTGGGACACAGCTGA 2026
733 sasprthlsrphenrclnlysnasnnglylvalvalgluaspasprpr 750
2027 TCAACGCGCATTTTAAAGGCGAAACTTCCAAATTTAAAGGCGGACGCG 2076
750 leasnargasnphenlysalatrthrthmetasnvalaThrlyllyasnalasr 766
```

```
2077 GTGGTTCC...CGCAATGTTGCCAAGTAAGCGGATGCGATTGGAC 2123
      :: ||| ||||| ||||| ::::: ||
767 LeuThySerGlyArgAsnValAlaAsnIleThrSerAsnIleThrAlaSer 783
2124 CAATCAGCGCCCAAGCAAGCTTTTGGTGTGCGACCGCATCAAGCGCACAA 2173
      ||||| ||||| ::::: |||
783 rAsnAsnAlaGluValHisIleGlyTrp.....LysThrGlyAspThrV 798
2174 TCTGTACAGCTTCGAGCTGCGGCTGTGCAATGTGTGCAAAAAC 2223
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||
798 aLysValAlaArgSerAspTrpThrGlyValThrCysHisAsnSerAsn 814
2224 ATTACCGCAGATAAGTGAATGCTTCATGACTAGACCGACATCAGCG 2273
      ::::: |||||::|||::|||::|||::|||::|||::|||
815 LeuSer...GluLysAlaLeuAsnSerPheAsnProThrAsnLeuArgI 830
2274 CAATGTCGATCTTCGCGGATCAGCTCATTTAAATCTCACAGGGCTGCCA 2323
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
830 yAsnValAsnLeuThrGluAsnAlaSerPheThrLeu...GlyLysAla 846
2324 CACTCAGCGCAATCTTAGTGCMAATGCGATACAGCTTATACAGTCAGC 2373
      ::||| |||||::|||::|||::|||::|||::|||::|||
846 snLeuPheGlyThrIleGlnSerIleGlyThrSerGlnValAsnLeuLys 862
2374 CACAAGCC.....ACCGAAAGCGCAACCTTAGCCTCGTGG 2411
      ::|||::|||::|||::|||::|||::|||::|||
863 GluAsnSerHisTrpHisLeuThrGlyAsnSerAsnValAsnGlnLeuAs 879
2412 C.....AATGCCCA..... 2421
      : ||||| |||||
879 nLeuThrAsnGlyHisIleHisLeuAsnAlaGlnAsnAlaAsnLysV 896
2422 ..GCAACATTATCAAGCCACATTAAAGCGCACATCGGCTTCGGC 2469
      ::|||::|||::|||::|||::|||::|||::|||
896 aLThrThrLysAsnThrLeuThrValAsn.....SerLeuSerGly 909
2470 AATGCTTCATT.....AATCTAAGCGACA 2495
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
910 AsnGlySerPheTrpTrpValAspPheThrAsnAsnLysSerAsnLys 926
2496 CGCCGCTACAAG.....GCGAGCTGACGCT..... 2523
      ::||| ||||| ||||| ||||| ||||| ||||| |||||
926 sValValValAsnLysSerAlaThrGlyAsnPheThrLeuGlnValAla 943
2524 .....TCC 2526
      : |||
943 sPlyThrGlyLupProAsnHisAsnGluLeuThrLeuPheAspAlaSer 959
2527 GGCACGCGTAAGCAACGTAAGCCATTCCGCACTCAACGGTAATGCTC 2576
      ::|||::|||::|||::|||::|||::|||::|||
960 AsnAlaThrArgAsnAsnLeuGluValThrLeuAlaAsnGlySerVal.. 975
2577 CCTAGCCGTAAGCAAGTATTCATTGAAAGCGCGCTTACCGGAC 2626
      ||||| ||||| ::::: |||
976 .....AspArgGlyAlaTrpLysLysLeuArgAsnValAsnGlyA 990
2627 AATCAGCGCGCGCAAGATACGCGATTACACTTAAAGACAGCGAATGG 2676
      ::||| ||||| ||||| ||||| ||||| ||||| |||||
990 rGlyTrpSP.....LeuTrpAsnProGlu... 997
2677 ACGCTGCGCGTAGCGCAAGTAATTAACCTTGAACCGCCAC 2726
      ::||| ||||| ||||| ||||| ||||| ||||| |||||
998 .....ValGluLysArgAsnGlnThrValAspThrThrAs 1009
2727 CATTCACACTATTCGCGCTATCCGACGATGCGGCGGCGCAACCG 2776
      ||||| ||||| ::||| ||||| ||||| ||||| ||||| |||||
1009 nIleThrThrProAsnAspIleGlnAlaAspAlaProSerAlaGlnSerA 1026
2777 GCAGT..... 2781
      ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1026 snAsnGluGluIleAlaArgValGluThrProValProProAlaPro 1042
2782 .....GCGACAGATGCGCGCGCGCGCTTCGCGCG 2813
```

```
1043 AlaThrGluSerAlaIleAlaSerGluGlnProGluThrArgProAlaG 1059
2814 TTCGCGCGCTCCATTATACGGTTACA..... 2841
      ::|||::|||::|||::|||::|||::|||::|||
1059 uThrAlaGlnProAlaMetGluThrAsnThrAlaAsnSerThrGluT 1076
2841 ..... 2841
1076 hAlaProLysSerAspThrAlaThrGlnThrGluAsnProAsnSerGlu 1092
2842 .....CCGCCACTTC 2852
1093 SerValProSerGluThrThrGluLysValAlaGluAsnProProGlnI 1109
2853 GGTAGAAATCC.....CGTTTCACACAGCTGACGG 2881
      : ||||| ||||| ||||| ||||| ||||| ||||| |||||
1109 uAsnGluThrValAlaLysAsnGluGlnGluAlaThrGluThrProThr 1126
2882 TAAAGCGCAATTG.....ACGGT 2901
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1126 lAsnGlyGluValAlaLysGluAspGlnProThrValGluAlaAsnThr 1142
2902 CAGGAACATTCGCTTATGTCGGAACCTTCGGCTACCGCAGCGACAA 2951
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1143 GlnThrAsnGluAlaThrGlnSerGlu.....GlyLysThrGluG 1156
2952 ATTGAAGCTGCGGAAGT..TCCGAAGCACTTACACTTGGCGGTCA 2998
      ::|||::|||::|||::|||::|||::|||::|||
1156 uThrGlnThrAlaGluThrLysSerGluProThrGluSerValThrValS 1173
2999 ACAATACCGCGCAAGAACT.....GCAACCTCGAACAA 3033
      ::||| ||||| ||||| ||||| ||||| ||||| |||||
1173 eArgL.....AsnGlnProGluLysThrValSerGlnSerThrGluAsp 1187
3034 TTGACGCTAGTGAAGAAAGAC..... 3057
      ::||| ||||| ||||| ||||| ||||| ||||| |||||
1188 LysValValValGluLysGluLysAlaLysValGluThrGluGluTh 1204
3057 ..... 3057
1204 rGlnLysAlaProGlnValThrSerLysGluProProLysGlnAlaGluP 1221
3057 ..... 3057
1221 roAlaProGluGluValProThrAspThrAsnAlaGluAlaGlnAla 1237
3057 ..... 3057
1238 LeuGlnGlnThrGlnProThrValAlaAlaAlaGluThrThrSerP 1254
3058 ....AACAAACCGCTGCGCAAAACCTTAATTCACCTCAAAAGCAAC 3103
      ::||| ||||| ||||| ||||| ||||| ||||| |||||
1254 oAsnSerLysProAlaGluGlu..ThrGlnGlnProSerGluLysThrAsn 1270
3104 .....ACGTGATGCGCGCGCTGCGGTACCAACATCATCGC 3141
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1271 AlaGluProValThrProValValSerGluAsnThrAlaThrGlnProTh 1287
3142 AAAGACGCGCAATTCGCGCTGCAATATCGGTCGTAAGAACAGAGCTTTC 3191
      ::||| ||||| ||||| ||||| ||||| ||||| |||||
1287 rGluThrGluGluThrAlaLysValGluLysGluLysThrGlnGluValP 1304
3192 CGACAAATCGCGCAAGCAGACAGCAAAACAGCGGGAAGAACACACAG 3241
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
1304 roGlnValAlaSerGlnGluSerProLysGlnGluGlnProAlaAlaLys 1320
3242 CGCAAAAGCTTGAAGCGCTGATTCGCGCGCGCGCGATCGCTGCAAAAG 3290
      ::||| ||||| ||||| ||||| ||||| ||||| |||||
1321 ProGlnAlaGlnThr.....LysProGlnAlaGluProAlaArgI 1334
3291 .....ACAGAAAGCTTCCGACACCGCGCGCGCAAGCA..... 3324
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

1334 uasnValLeuThrThrLysAsnValGlyGluProGlnProGlnAlaGlnP 1351
3324 ..... 3324
1351 roGlnThrGlnSerThrAlaValProThrThrGlyGluThrAlaAlaAsn 1367
3324 ..... 3324
1368 SerLysProAlaAlaLysProGlnAlaGlnAlaLysProGlnThrGluPr 1384
3325 ..GCGGGGAAATGTCGGCATTATGACGGCGAGAA..... 3360
::: |||||::: ::::::::::::::|
1384 oAlaArgLysAsnValSerThrValAsnThrLysGluProGlnSerGlnT 1401
3360 ..... 3360
1401 hrSerAlaThrValSerThrGluGlnProAlaLysGluThrSerSerAsn 1417
3360 ..... 3360
1418 ValGluGlnProAlaProGlnAsnSerLLeasnThrGlySerAlaThrTh 1434
3360 ..... 3360
1434 rMetThrGluThrAlaGluLysSerAspLysProGlnMetGluThrValT 1451
3360 ..... 3360
1451 hrGluAsnAspArgGlnProGluAlaAsnThrValAlaAspAsnSerVal 1467
3360 ..... 3360
1468 AlaAsnAsnSerGlnSerSerGlnSerSerArgArgArgSerVal 1484
3361 ...GAGAAAAACGGGTGACGGCGGATPAAAGACACCGCTTGGCGAAGC 3406
::: |||||::: ::::::::::::::|
1484 lSerGlnProLysGluThrSerAlaGluGluThrThrValAlaSerThrG 1501
3407 AGCGGGAAGCGGAACCGGGCGGTACACCGCTTCCCGGGCGCGC 3456
||::: ::: |||||:::
1501 lngluthrThrValAspAsnSerValSerThrProLysProArgSerThr 1517
3457 CCGCGCCGCGGGATTTGGCCGACACTGCACCCCAACCGAGCCCAAC 3506
||::: |||||::: ::::::::::::::|
1518 ArgThrArgArgSerValGlnThrAsnSerThrGluProValGluLeuPr 1534
3507 GCAG..... 3510
1534 oThrGluAsnAlaGluAsnAlaGluAsnValGlnSerGlyAsnAsnValA 1551
3511 .....CGGACCTGATCAGCGCTTATGCCAATAGC 3540
1551 lAsnSerGlnProAlaLeuAlaGlnAsnLeuThrSerLysAsnThrsmLa 1567
3541 GGTTCAGTGAATTTCCGCC.....ACGCTCACAGCGT 3575
||::: |||||::: ::::::::::::::|
1568 ValLeuSerAsnAlaMetAlaLysAlaGlnPheValAlaLeuAsnValG 1584
3576 TTTTCGCCGTCAGGAGGAGATTAGCCGGCTA...TTTGGCGAAGCCGC 3622
||::: |||||::: ::::::::::::::|
1584 yLysAlaValSerGlnHisLLeSerGlnLeuMetAsnAsnGluGlyG 1601
3623 GCAACGCCGTTTGACACAGCGGATCCGGACACCAACACTACCGTTGC 3672
::: |||||::: ||::: |||||:::
1601 lntyrAsnValTrpLLeSerAsnThrSerMetAsnLysAsnTrpSerSer 1617
3673 CAAGATTTCGGCGCTTACCGCCACAAACCGACTCGCCCAATGGGAT 3722
::: |||||::: ::::::::::::::|
1618 GluGlnTyrArgArgPheSerSerLysSerThrGlnThrGluLeuGlyTr 1634
3723 GCGAAGAAACCTCGGACGCGG...CGCGTCGATCCGTTTTCGCAAC 3769
::: |||||::: ::::::::::::::|
1634 pAspGlnThrLLeSerAsnAsnValGlnLeuGlyGlyValPheThrTyrV 1651

```

```

3770 ACCGGACGAAACACCTTCGACGAGCGCATCGCACTCGGACGGCT 3819
||::: |||||::: ||::: |||||:::
1651 aLArgAsnSerAsnAsnPheAspLysAlaSerSerLysAsnThr...Leu 1666
3820 GCCCAGCGGCGGCTTTCGGGCATACCGCATCGAC...AGGTTCATAT 3866
||::: |||||::: ||::: |||||:::
1667 AlaGlnValAsnPheTyrSerLysTyrAlaAspAsnHisTrpTyrLe 1683
3867 CCGCATCAGCGCGCGGCTTTCAGCGGCGGACGCGCATTCAGCGCA 3916
||::: |||||::: ||::: |||||:::
1683 uGlyLLeAspLeuGlyTyrGlyLysPheGlnSerAsnLeuGlnThrAsn 1700
3917 TCGGAGCAAAATCCCGCCGCGGCTTCATTCAGCATTCAGCGCA 3966
::: |||||::: ||::: |||||:::
1700 snAsnAlaLysPheAlaArgHisThrAlaGlnLeuGlyLeuThrAlaGly 1716
3967 TACCGCGCGGCTTTCGGCGATTCGCGATCGAAGCGGACATCGCGCAAC 4016
::: |||||::: ||::: |||||:::
1717 LysAlaPheAsnLeuGlyAsnPheAlaValLysProThrValGlyValAr 1733
4017 GCGCTATTTCGTCGCAAAAGCGGATTCAGCGCTACGAAACGTCAATATCG 4066
::: |||||::: ||::: |||||:::
1733 gTyrSerTyrLeuSerAsnAlaAspPheAlaLeuAlaGlnAspArgLLe 1750
4067 CCACCCCGCGGCTTCATTCACCGCTACCGCGCGGCAATAGCGCAT 4116
::: |||||::: ||::: |||||:::
1750 yValAsnProLLeSerValLysThrAlaPheAlaGlnValAspLeuSer 1766
4117 TATTCATTCAAACCGCGGCAACACATTCATTCACCGCTTATTCAGCT 4166
||::: |||||::: ||::: |||||:::
1767 TyrThrTyrHisLeuGlyGlu...PheSerLLeThrProLLeuSerAl 1782
4167 GTCTTATTCGATCCGCTTCGGGCAAGTCCGACACCGCGTCAATACG 4216
||::: |||||::: ||::: |||||:::
1782 aArgTyr...AspAlaAsnGlnGlyAsnGlyLysLLeAsnValSerValT 1798
4217 CCGTATTCGCTCAGATTCGCGCAAAACCGCATCGGATAGGGCGTA 4266
||::: |||||::: ||::: |||||:::
1798 yAspPheAlaTyrAsnValGluAsnGlnGlnGlnTyrAsnAlaGlyLeu 1814
4267 AACGCCGAATCAAGTTTCACGCTGCCCTCAGCGCGCGCCGCA 4316
||::: |||||::: ||::: |||||:::
1815 LysLeuLysTyrHisAsnValLysLeuSerLeuLLeGlyLeuThrL 1831
4317 AGGCGCCGACACTGGAAGCGGACACAGCGCGGCTCAATAGGCTAC 4365
||::: |||||::: ||::: |||||:::
1831 sAlaLysGlnAlaGluLysGlnLysThrAlaGluValLysLeuSerPhe 1847

seq_name: SwissProt_40:YPIA_ECOLI
seq_documentation_block:
ID YPIA_ECOLI STANDARD; PRT; 1569 AA.
AC P52143; P76610; P77017; P77019;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical outer membrane protein YPIA.
GN YPIA OR B2647.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).

```


[2]
 SEQUENCE FROM N.A.
 RP STRAIN-K12;
 RC MEDLINE=97349980; PubMed=9205837;
 RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
 RA Mizobuchi K., Mori H., Nakase S., Nakamura Y., Nishimoto H.,
 RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubaram S.,
 RA Tagami H., Takahashi H., Takeda U., Takemoto K., Uehara K., Wada C.,
 RA Yamagata S., Horuchi T.;
 RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
 RT - K12 genome corresponding to 50.0-66.8 min on the linkage map and
 RT analysis of its sequence features.";
 RL DNA Res. 4:91-113(1997).
 CC -1- SUBCELLULAR LOCATION: Outer membrane (Potential).
 CC -1- SIMILARITY: STRONG, TO BORDETELLA PERTUSSIS.
 CC
 CC
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U06840; AAA79815.1; ALT SEQ.
 DR EMBL; AE000350; AAC75695.1; -
 DR EMBL; D90889; BAA16514.1; ALT INT.
 DR EMBL; D90890; BAA16518.1; ALT INT.
 CC Ecogene; EG13213; ypja.
 KW Hypothetical protein; Outer membrane; Complete proteome.
 SO SEQUENCE 1569 AA: 162774 MW: 0D006A9A3154596 CRC64;

```
alignment_scores:
  Quality: 364.50
  Ratio: 0.485
  Percent Similarity: 46.022
  Length: 1634
  Gaps: 81
  Percent Identity: 19.951
```

alignment_block:

US-09-303-518D-649 x YPJA_ECOLI ...

Align seg 1/1 to: YPJA_ECOLI from: 1 to: 1569

```

445 GGGAGTAAGGACCATCCCTTATGGGGGAGATTATCATATGCCGCTTTGCA 494
||||| ..... : : : : :
97 G|YthrlrlnAsnAsnTlrlrhlleAsnHlselYglInAsnValHl 113
495 TAAATTTGTCACAGATGAGAACCTGTTGAATATACCAAGTTATATGAT. 543
| ||| : : : : : : : : : : :
113 sG|YglYValSerlnsnglSerleuileGlnSerGlyglYTrGlnAsprl 130
544 ..... GGGCGAAATATATTCGATCAAAATATATTCCTGACCGGT 585
: : : : : : : : : : :
130 leGlySerHlsAsnAsnPhenValG|YglInAlAsn.....AsnThr 143
586 CGATTTGGGGGAGCGACGACCAATATTTGGCGATCGATCAAGTACGACCA 635
||||| : : : : :
144 ThrleAsnnglYglYatrgIn..... 150
636 TAACCGCGAAAGTTCAATATCATATTTGCA.....AGTGCCTATCTTGGC 679
||| ||| : : : : :
151 .....SerleHlsAspGlyglYleSerThnglYthrlThrl 163
680 TCgTTGCTGGCATACTTGTTCACAAATATGA.....TCAGTGGGTGGC 723
: : : : : : : : : : :
163 leGlnSerG|YglAsnGlnAspValYtlYsnglYlYleSerAsnnglYThr 179
724 ACACTCACTTAGTACTGAAAAATATTA...CATAGCCCATATGGTTT 770
||| : : : : : |||
180 ThrleYsnglYglYalaserArGvalG|YglYSerAlAsnnglYtl 196
771 TTATACCAACAGAGGCTCA..... 789

```

```

196 eleuileasprclgysgerglnilevalylsvalglnlglnihlsalasp 213
    :||| ||||| |||
210 ..... TTGGGACAGAGGGCTCCACCAATGTTATCTATGATGCCAA 831
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
213 lylrthrlleasnlyserclyserslnsvalvalglnlglyserleu 229
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
832 AAGCAAAAGTGGTAAATTAATGGGCTATTCGCAAGGCAACCCATAT 681
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
230 Alathrsnplrthrlleasnlgly..... GllAraglplutya 242
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
882 AGGA..... AAAACAAATGGCTTCAGC 904
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
242 lglulnserthvalglulrthrlthrllellysasnlglyglulnla 259
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
905 TGGTTCGAAAGATGGCTGTATGATGAAATCTTGTGCGAGATACCAT 954
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
259 rglvalylrgluserarglaleuasnplrthrlleclulclgluln 275
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
955 TCAGATATTTCAGCAACCACTCAAAATGGCAAAATCTCTTTAACGAC.. 1007
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
276 serleuasnserlyserthrllalylasnthrnlslletylsercl 292
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1003 ..... GATTAATATGGCAGAGAAATCAATGCCAAACATG 1033
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
292 ythrllelvalylspsnrlserthserasvallelgulvalylus 309
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1040 AACCAATTCCTCGCTAAATAGA..... 1066
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
309 erglglulvalleuasnprvalrglglulrthrlalathrsnvalthrln 325
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1063 ..... TTAACACAGACCCGTCAATTGTTATGTTTC 1097
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
326 hisaprlalalleuylserthrsnplr..... Asnlglythrth 339
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1098 TTTATCCGACAGCAAGAAAGACCTGTAT..... CATGCG 1133
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
339 rvalserglythrlnserglulalatheserllelnlsnhtsval 356
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1136 CA..... GGTGCT..... GTCCACAGTAT 1155
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
356 laasprsnvalleuaglualsnlglylnlsleuasnrlleasnrlatyr 372
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1155 ..... 1155S
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
373 glyseralasnlysrthrllellysasprlysglythrmetserval 389
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1156 ..... CGACCCAGATCAATATAGA..... 1176
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
389 uthrlnalalylaspralathrarglleasnplclglualmet 406
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1177 ..... GAAAT 1182
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
406 sprvalalaglysnalathrsnplrlelleasnlglylthrclnln 422
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1183 ATTTCCTTATTCAGCAAGAAAGCAATGATACTTACCACACAT 1232
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
423 llesasnlyrglylealathrgly..... Thrlnll 434
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1233 CAATCAAGT..... GCTGAGGATTAATTTCCAGAG 1267
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
434 easnserglythrlnasnlllelyserclgly..... Lysalaa 448
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1268 ATTTCAGCTCCGCTGAAATAC..... GAAACTGG 1302
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
448 sprhrthrllelesergerlyserarglnvalvalglulysasprly 464
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1303 CAAGGCGCGGCTCATATCAGTAGACAGTACCCGTACTTGGAAGT 1352
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
465 thrAlaleglyserasnleaserlalelylserleuilevalyltrth 481
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1353 AAACGCGGTGCAACGAC..... CGCGTGT 1378
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

481 rgllygllylealhisgllyvalasnnglnlthrlyseralaleuvala 498
 1379 CCAAAATCGCAAGCAGC.....CTGCAC 1404
 498 laasnhrlygllythrleasplleugllytrlyasnlyleuserhis 514
 1405 GTTCACCAAGG.....GAAACCAAG 1430
 515 Phehrlethrlygllylualasnlyvalalleuglnlunhrly 531
 1431 CTGATCGAGTGGGAGCGATTCATTTGATCGACAGCAGCAGC 1480
 531 yglulethrlyvalalalythrseralalyasnthrlyleaspt 548
 1481 ATAAAGCAAA.....AAACACCTTATGTAATCGCCTTG 1518
 548 hrlygllyllyleuilevalgllylualalythrleaspt 564
 1519 GTACGGCGAGGGGTACGTCACATGATCCGAT..... 1554
 565 leuasnasnlygllyvalleugllylualaspolygllylualalyshl 561
 1555AATCATGTCACCCCGCAAACTCTATTGCGCTTCGCGCGGAC 1600
 581 svaiglnlgnlserlygllylaleuileasethrlyserly 598
 1601 GTTTG.....GATTTAAGCGGCGATTCGCTTCACCGTATTCAA 1644
 598 hrleuilegllylthrlyasnserlygllylaspalaphetyr...learg 613
 1645 AATCCGATGAGGGGATGATGTCACACCAATCAAGCAAAAGATC 1694
 614 asnergllylalyasnvalleu.....gluasnalyse 627
 1695 CACCGTTACATTCACAGCAATAAGATATTGTACACCGCAATTA 1744
 627 rleuthrlyvalthrlyserargalalyaspthrlyleleasnlaa 644
 1745 ACAGCTTGATAGCAAAAGAAATTCGCTCAACGCTGGTGGGAG 1794
 644 sn.....gllylmetaspal.....tyrly... 651
 1795 AAAGT.....ACGACCAAAAGCAGCGCGCTCAACCTTTGTTACA 1838
 652 lysasprvalythrlyleuasnseralaglythrlnhrlyetlyal 668
 1839 GCCCGCCAGAGACCGCCTGCTGCTTCGCGGGAACAATTTAA 1888
 668 aseralathrserasplysalaasnlylelygllylsglnthrlyalt 685
 1889 ACGCAATCATCAGCAACCAAC.....GGCAACTGTTTTTCAGC 1929
 685 yrglyleualathrlylualasnlyleugllylgnlnlevalasp 701
 1930 GGCAGACCAACACCGCAGCCTACATCTTAAACGAC..... 1968
 702 gllyllyserthrlylly.....Thrhisleasnlygllylthrglnth 716
 1969C 1969
 716 rvalglasnlyrlyllyllylaleasnthrlyaspillevalserlyleug 733
 1970 ATTGGTCGCAAAAGAGGATTCCTCGGGAATCGTG...TGGAC 2016
 733 lnglnllemetalaasnlythrlylagnlylserlleleasnlylly 749
 2017 AACGACTGATCAACCGCATTTAAAGCGAAACTTCCAAATTTAAG 2066
 750 sergllylvalasnlygllyllyleualagnlyllyasnservalleuasn 766
 2067 CGACAGCGGGGTGTTCCCGCAATGTTCCAAAGTGAAGCGGATTTGGC 2116
 766 pelygllylthrlyleuaspvalarglyllyllyseralathrgly..... 780

2117 ATTGAGCAATCAGCCCAAGCAGTTTGGTGGCAGCGCATCAAGC 2166
 781 ..lleglnlserserlygllylaleuvalalathrlyaralathr 796
 2167 CACACATCTGTACAGCTTCGAGCTGAGCGGTCTGACAAATGTGTGCA 2216
 797 argvalthrlylthrlyargalasp.....glylvalalapheserilegl 811
 2217 AAAA.....ACA 2224
 811 ugllyllylaleasnlnleleuvalaleasnlygllyllyleuvalthr 828
 2225 TTACGAGAT..... 2235
 828 algllylseraspthrlyseraspthrlylvalasnmetglylyarg 844
 2236AAGTATTCCTTCATTCATTAAGCAGCATCAGCGCA 2276
 845 gllyllyllylthrlylathrlylathrglylthrlylthrlyllyl 861
 2277 TGTGATCTT.....GCCATCAGCGCTCATTTAATCTCAGCAG 2314
 861 ygllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 878
 2315 GCTTGCACACTCAACGCAATCTTATGCAATGGCAT..... 2355
 878 lytlelnthr.....valseralaleasnlygllyllyllyllyl 890
 2356 ACAGCTTATACAGTACGACCAAGCCCAAGCAAGCGCAGC..... 2397
 891 thrlyllyllyllyllyllyllyllyllyllyllyllyllyllyl 907
 2398 ..CTTACCTCGTGGCAATGCCAGCAACATTTATCAAGCCATTA 2446
 907 athraspllevalgllyllyllyllyllyllyllyllyllyllyl 924
 2447 ACGCAACATCGGCTTCGCGCAT.....GCTTCAATTAATCA 2487
 924 snly...lleglnlserlyllyllyllyllyllyllyllyllyllyl 939
 2488 ACGGACAGCGCTTACAAACGCGAGTTCAGCGCTTCGCGCACTAA 2537
 940 sergllyllyllyllyllyllyllyllyllyllyllyllyllyl 956
 2538 GGCAAACGTAAGCATTCGCACTCAACGTAATGCTCCCTAGCGATA 2587
 956 uvalleualaglylthrlylaleargaspserthrlyllyllyllyl 972
 2588 AGCGATATTCATTTGAAGCAGCGCTTACCGGCAAAATCAGCGC 2637
 973 ..Alametgllyllyllyllyllyllyllyllyllyllyllyllyl 988
 2638 GGCAGATTCAGCGCATTCATTAAGCAGCGATTCAGCGCTGCTC 2687
 989 llyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 1005
 2688 AGCGACGCAATTA..... 2700
 1005 gyllyllyllyllyllyllyllyllyllyllyllyllyllyllyl 1022
 2701GGCAATTAACCTT... 2715
 1022 hrleualaspaleaservalserlyllyllyllyllyllyllyllyl 1038
 2716GACACGCGCACC...ATTACATCAATTCGCTATGCG... 2751
 1039 Thrlyllyllyllyllyllyllyllyllyllyllyllyllyllyl 1055
 2752 ..CAGATTCGCGCAGCGCGCAACCGCGCTGCGACAGATGCGCGCGC 2800
 1055 ethraspseralathrlylthrlyllyllyllyllyllyllyllyl 1072

```
2801 GCCGTTCCGCCGTTCCGCCGCT..... 2823
      ::::|
1072 laAspLeuThrAlaIaSerArgIySerValTrpLeuAsnSerAsn 1088
2824 .....TCCCTATTAA.. 2832
      |||||
1089 SerCysAlaGlyThrSerAsnCysGluTyrArgValAsnSerLeuLeu 1105
2833 .....TCCGTTACACCGCCGCACTTCGG 2854
      ::::|
1105 uAsnAspGlyAspValTyrLeuSerAlaGluThrAlaIaProIaThrT 1122
2855 TAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAATTGAACGGTACG 2904
      ::::|
1122 hrAsnGlyIleTyrAsnThrLeuThrAsn...GluLeuSerGlySer 1137
2905 GGACATTTCCGCTTATGTCGGAACTCTTGGCTACCGCAGCGCAAAATT 2954
      |||||
1138 GlYAsnPhetYrLeuHisThrAsnValAlaGlySerArgIyAspGlnLe 1154
2955 GAACGTTGGCGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACATA 3004
      ::::|
1154 uValValAsnAsnAsnAlaThrGlyAsnPhetYrIlePhetValGlnspt 1171
3005 CCGGCAACGAACTCGCAAGCTCGAACATTTGACGGTAGGAAGGAA 3054
      |||||
1171 hrGlyValSerProGlnSerAspAspAlaMetThrLeuVal..... 1184
3055 GACACAAACCGCTGTCCGAAACCTTAATTCAACCTCGCAAAACGAA.. 3102
      |||
1185 .....LysThrGlyIyGlyAspAlaSerPhetThrLeuGlyAsnThrG 1199
3103 ....CACGTGATCCGCGCGCTGAGCGCTTACCAATCATCCGCAAGACG 3148
      |||||
1199 yGlyPhetValAspLeuGlyThrTyrGluTyr...ValLeuYsSerAsp 1215
3149 GCGAGTCCGCCCTGCATATCCGCTCAAGAACAGAGCTTCCGACAA 3198
      |||
1215 Ly.AsnSer..... 1217
3199 CTCGGCAGGCGAGAACCAAAAACAGGCGGAAAGAACAGCGCGCAAG 3248
1218 .....AsnTTPA 1220
3249 CCTTGACGGCTGATTGCGCGCGCGCGATGCCGTGCAAGAAACAGAA 3298
      |||||
1220 snLeuThrAsnAspValLysProAsnProAspPro..... 1231
3299 GCGTTCGCAACCGCGCGCGGAGCGAGCGGGGAAATCTCGGCATTATG 3348
      ::::|
1232 ...IleProAsn..... 1234
3349 CAGGCGGAGAGAAAGAAAAACGGGTGACGGGATAAAGACACCGCTT 3398
1234 ..... 1234
3399 GCGGAAACAGCGGAAAGGAAACCGCGGCTACACCGCTTCCGCC 3448
1235 .....ProLysProAspProLysProAspProLysPro 1245
3449 GCGCCCGCGCGCGCGCGGATTGTCGCAATCAACCCCAACCGGAG 3498
      |||
1246 AspPro.AsnProLysProAsp...ProThrProAspProThrProthr 1260
3499 GCGCAACCGCAGCGGACCTGATCAGCGGTATGCAATAGCGGTTTACG 3548
      |||
1261 ProValProGluLysArgIleThrPro.....SerThrAlaIaValLe 1275
3549 TGAATTTTCCGCGCTCAACAGCGCTTTC.....GGCG 3583
      ::::|
1275 uAsnMetAlaIaThrLeuProLeuValAlaPheAspAlaGluLeuAsnSer 1292
3584 TACAGACGAATTAGACCGCGTATTGCGCAAGACCGCGCAACGCGCTT 3633
```

```
1292 leArgIleuArgLeuAsnIleMetLysAlaSerProHisAsnAsnVal 1308
      ::::|
3634 TGCACAGCGCGCATCCGGACACCAACACTACCGTTCCGCAAGATTCCG 3683
      |||
1309 Trp...GlyAlaThrTyrAsnThrArgAsnAsnValThrThAspAlaG 1324
3684 CGCTTACCGCCCAACAAACCGACCTGGCCAAATCGTATGCAAAAAAC 3733
      ::::|
1324 yAlaGlyPheGluGlnThr.....LeuThrGlyMetThrValGlyI 1338
3734 TC.....GGCAGCGCGCGCTGCGCTCTG 3759
1338 leAspSerArgAsnAspIleProGluGlyIleThrThrLeuGlyAlaPhe 1354
3760 TTTTGCACACACCGCAGCAACACCTTCGACAGCGC..... 3798
1355 MetGlyTyrSerHisSerHisIleGlyPheAspArgIyGlyHisGlySe 1371
3799 .....ATCGGCACTCGGACAGCGCTTCCGACAGCGCGG 3832
1371 rValGlySerTyrSerLeuGlyGlyTyrAlaSerTrpGluHisGluSerG 1388
3833 TTTTC.....GGCAATACGGCATCGACAGGCTC.....TAC 3864
1388 LysPhetYrLeuAspGlyValValLysLeuAsnArgPheLysSerAsnVal 1404
3865 ATCGCAATCAGCGCGCGCGGCTTTTAGCAGCGCGACGCTT...TCGA 3911
      |||
1405 AlaGlyLysMetSerSerGlyGlyAlaAlaLysSerTyrHisSerAs 1421
3912 GCGCATCGGAGCAAAATCCGCGCGCGCTGCTGCATTACGGCATTCAG 3961
      ::::|
1421 nGlyLeuGlyLysHis.....IleGluT 1429
3962 CACGATACCGCGCGGCTTCCGCGGATTCGGCATCGACCGCAATCGCGC 4011
      ::::|
1429 hrGlyMetArgPheThrAspGlyAsnTrpAsnLeuThrProTyrAlaSer 1445
4012 GCACGCGCTATTTCGTCCAAAAGCGGATTACGCTACGAAAC..... 4056
      |||
1446 LeuThrGlyPheThrAlaAspAsnProGluTyrHisLeuSerAsnGlyMe 1462
4057 .....GTCAATATCGCCACCGCGGCTTGCAATCAACCGCTACCGC. 4098
1462 LysSerLysSerValAspThrArgSerIle.....TyrArg 1475
4099 ..GCGGCAATTAAGCAGCATTTATTCATTCAAACCGCGCAACACATTCC 4146
      |||
1475 IuLeuGlyAlaThrLeuSerTyrAsnMetArgLeuGlyAsnGlyMetGlu 1491
4147 ATCAAGCCTTATTGAGCGCTGCTCATACGATCCGCTTCGGCAAAAGT 4196
      ::::|
1492 ValGluProThrLeuLysAlaAlaValArgLysIuPheValAspAsp 1508
4197 CCGAACACGCGTCATACCGCGGCTATGCTCAGAGATTTC.....GGCA 4240
      |||
1508 nArgValLysValAsnSerAspGlyAsnPhetValAsnTyrLeuSerGly 1525
4241 AAACCCGACAGTCGGAATGGGCGCTAAACGCGCAATCAAGGTTTCACG 4290
      ::::|
1525 rGArgGlyIleTyrGlnAlaGlyIleLysAlaSerPheSerSer...Thr 1540
4291 CTGTCCCTCCACGCGTCCGCGCGCAAGGCGCGCACTGCAACGCAAC 4340
      |||||
1541 LeuSerLysHisLeuGlyValGly.....TyrSerHis 1551
4341 CAGCGCGGCAATCAA.....TTAGCTTACCGCTGCG 4371
      |||
1551 sSerAlaGlyValGluSerProTrpAsnAlaValAlaGlyValAsnTrp 1567
```

seq_name: SwissProt_40:PERT_BORBR


```

279 GlyProLeuLeuAspGlyTyrGlyVal..... 288
2223 CATTACGACGATAAAGTATGCTTCATTGACTAGACGACATCAGC. 2271
289 .....AspValSerA 292
2272 ..GGCAATGTCATCTTGGCGATCAGCTCATTTAAATCTCACAGGGCTT 2319
292 sPserThrValAspLeuAlaGln.....SerLeuValGluAla 304
2320 GCCACACTACAGCGCAATCTTAAGCAATGGCGATACAGCTTATACGT 2369
305 ProGlnLeuGlyAlaAlaIleArgAlaGlyArgGlyAlaArgValThrVal 321
2370 CAGCCACACAC.....GCCACCCAAAAGCGCAACCTTACCTCGTGGGCA 2413
321 lSerGlyGlySerLeuSerIleAlaProHisGlyAsnValIleGluThrGlyG 338
2414 ATGCCCAAGCAACATTTAAT.....CAAGCCACATTA 2445
338 lYcIylAlaArgArgPheProProAlaSerProLeuSerIleThrIleu 354
2446 AAGCGAACAACATCGGCTTCGGCAATGCT...TCATTTAATCTAAGCA 2492
355 GlnAlaGlyAlaArgAlaGlnGlyArgAlaLeuIleuTyrArgValLeuPr 371
2493 CCAGCGCTACAAAACGGCAGTGTGACGCTTCGGCAACGCTAAGGCA 2542
371 ogIuProValLys.....LeuThrIleuAlaGlyAlaGlnGly. 384
2543 ACGTAAACCATTCGCGACTCAACGGTATGCTCCCTACGCGATAAGCA 2592
385 .....GlnGlyAspIleValAlaThr..... 391
2593 GATTTCATTGTAAGACGCGCTTACCGGCAATCAGCGCGGCA 2642
392 .....GluLeuProProIleProGlyAlaSerSerGlyProle 404
2643 GGATACGCGATTACACTTAAAGACAGCAATGACGCTGCGTACGCA 2692
404 uAspValAlaIleu...AlaSerGlnAlaIleArgTrpThr...GlyAlaThrA 419
2693 CGGATTAACGCAATTTAAACCTTGACACAGCGCATTAACCTCAATTC 2742
419 rGAlaValAspSerLeuSerIleAspAsnAlaThrTyrValMet..... 433
2743 GCCTATCGCCAGATGCGGCGACGAGGCGCAACCGGCGATGCGACGATGC 2792
434 .....ThrAsp... 435
2793 GCGCGCGCGCGCTTCGCGCGCTTCGCCCTATATTCGTTAC 2842
436 .....AsnSerAsnValGlyAlaIleuArgLeuAlaAs 446
2843 CGCCAACTTCGTAGAA.....TCCCGTTTCAC 2871
446 eAspGlySerValAspPheGlnGlnProAlaGluAlaGlyArgPheLys 462
2872 ACGCTGACGCTTAACGCAATTAAGGCTGAGGACATCCGCTTAT 2921
463 CysLeuMetValAsp...ThrLeuAlaGlySerGlyLeuPheArgMetAs 478
2922 GTCCGAACCTTCGGCTACCGCAGCAAAATTGAAGCTGGCGGAAGTT 2971
478 nValPheAlaAspLeuGlyLeuSerAspLysLeuValValMetArgAspA 495
2972 CGGAAGGACACTTACACCTTGGCGTACCAATACCGGCAAGCACTGCA 3021
495 lAspGlyGlnHisArgLeuLeuValArgAsnSerGlySerIleProAla 511
3022 AGCCTCGAACAATGACGCTAGTAGAAGAAAAGACAAACAACGCGCTGC 3071
512 SerGlyAsnThrMetLeuLeuValGln.....ThrProArgGly 524
3072 CGAAACCTTAATTCACCTGCAGAAACGAA.....CACGTGCATGCCG 3115
524 ySerAlaAlaThrPheThrIleuAlaAsnLysAspGlyLysValAspIleG 541
3116 GCGCGTGGCGTTTACCACTCATCCGCAAAAGACGGGAGTTCCGCTGCAT 3165
541 lYThrTyrArgTyrArgLeu..... 547
3166 AATCCGTTCAAGAAACAGACGCTTCGACAACTCGGCAAGCAAGCAGC 3215
547 ..... 547
3216 CAAAAACAGCGGCAAAAGACAGCGCAACGCTTACGCGCTGATTG 3265
548 .....AlaAlaAsnGlyAsnGlyIntPr.....SerLeuValG 559
3266 CGCGCGGCGGATGCCGTGCAAAAAGACAGAAACGTTGCCGACGCGGC 3315
559 lYAlaLysAlaProProAlaPro.....LysPr 568
3316 CGCGACGACGCGGCGGAAATGTCGCATTATGACAGCGGAGAAAGAA 3365
568 oAlaProGlnProGlyPro..... 574
3366 AAACGGGTGACAGCGCGATTAAAGACACCGCTTGCCAAACAGCGCAAG 3415
574 ..... 574
3416 CGAAACCGCGCGCGCTAACACCGCTTCGCCCGCGCGCGCGCGCG 3465
575 ...GlnProGlyProGlnProProGlnProProGlnProProGlnPro 590
3466 CGGATTTTCCGCAACGCAACCCCAACCGGACCGCAACCGCAAGCGCA 3515
590 roGlnArgGlnProGluAlaLProAlaProGlnProProAlaGlyArgGly 606
3516 CCGTATGACCGCTTATGCAATACCGGTTGAGTGATTTTCCGCCACGC 3565
606 uIleu...SerAlaAlaAlaAsnAlaAlaValAlaThrGlyGlyValGlyL 622
3566 TCAACAGCGTT.....TTCGCCGTACAGACGCAATTAACCGCGATT 3609
622 euAlaSerThrLeuTrpTyrAlaGlnIleSerAsnAlaLeuSerLysArgLeu 638
3610 GCCGAACGCGCGCAAC.....GCGCTTTGCAACAGCGCAT 3647
639 GlyGluLeuArgLeuAsnProAspAlaGlyGlyAlaThrPheLysArgLph 655
3648 CCGGAGACCAACAACACTACCGTTGCAAGATTCGCCGCTACCGGCAAC 3697
655 eAlaGlnArgGlnGlnIleuAspAsnArgAlaGlyArgArgPheAspIle 672
3698 AAACCGACTGCGCCCAATCGGTATGCAAAAACCTC..... 3735
672 ySValAlaGlyPheGluLeuGlyAlaAspHisAlaValAlaValAlaGly 688
3736 GCGACGCGCGCGCTGCGCATCTGTTTCCACACCGGACCGCAAAACAC 3785
689 GlyArgTyrPheIleuGlyGlyLeuAlaGlyTyrThrArgLysAspArgL 705
3786 CTTC...GAGACGCGCATCGCAACTCGGCAACGCTGCCACGCGCGCG 3832
705 yPheThrGlyAspGlyGlyGlyHisThrAspSerValHisValGlyGly 722
3833 TTTTTCGCGCAATACCGCATGACAGTTCTTACATCGGCATACGCGGCG 3882
722 yAlaThrTyrIleAlaAsnSerGlyPheTyrLeuAspAlaThrLeuArg 738
3883 GCGGCTTTAGCAGCGGACGCTT.....TCAGACGC..... 3915
739 AlaSerArgLeuGlnAsnAspPheLysValAlaGlySerAspLysTyrAl 755

```


130ValSerAspThrIArg 135
1808 AACGACGAGGCGGTCAACCTGTTTACCAGCCGCC 1845
135 SPASpAspGlyIleAlaLeuTYrValAlaGlyGlnAlaGlnAlaSer 151
1846 GCAGAGACCGCACCTGCTGCTTTCCGGGAAACAATTAAAA 1889
152 IleAlaSpSerThrLeuGlnGlyAlaGlyIleValArgValGlnArgGly 168
1890 CGGACATCATCGCAACAAACGCAACTGTTTTCAGCGGC 1932
168 yAlaAsnValThrValGlnArgSerThrIleValAspGlyGlyLeuHis 185
1933AGACCAACACCGCAGCCCTACATCATTTAAAGCAACAT 1971
185 IeGlyThrLeuGlnProLeu 191
1972 TGGTCGCAAAAAGAGGCAATTCCTCGGGGAAATCGTGGGACAACGA 2021
192GlnProGlnAspLeuProProSerArgValValLeuGlyAspTh 206
2022 CTGGATCAACCGCACAT 2038
206 rSerValThrAlaValProAlaSerGlyAlaProAlaAlaValPheValP 223
2039 TTAAGCGGAAACTTCACAAATTAAAGCGGACAGCGGCTGTTCCCGC 2088
223 heGlyAlaAsnGlnLeuThrValAspGlyGlyHisIleThrGlyGlyArg 229
2089 AATGGTCCCAAGTGAAGAGCGATTGGCAATTGACATCAGCCCAAGC 2138
240 AlaAlaGlyValAlaAla 245
2139 AGTTTGTGTGTGCAACCGCATCAAGCCACACATCTGTACCGTTCCG 2188
245 aMetAspGlyAlaIleValHisLeuGlnArgAlaThrIleArgArgGlyA 262
2189 AC 2190
262 spAlaProAlaGlyGlyAlaValProGlyGlyAlaValProGlyGlyAla 278
2191TGGACGGGTCTGCAAA 2207
279 ValProGlyGlyPheGlyProLeuLeuAspGlyTyrGlyVal 293
2208 TTGTGTGCAAAAACCATTAACGAGATTAAGTATTGCTTATGACTA 2257
293 293
2258 AGACGACATCAGC...GGCAATGTGATCTTCCGATCAGCGTCATTTTA 2304
294AspAlaSerAspSerThrValAspLeuAlaGln 304
2305 AATCTCACAGGCGCTTCCACACTCAACGCGCAATCTTAGTCAAAATGGCGA 2354
305 SerIleValGlnAlaProGlnLeuGlyAlaIleAlaIleArgAlaGlyArgGly 321
2355 TACAGCTTTACAGTCAGCCACAC...GCCACCCAAAAGCGCAAC 2398
321 yAlaArgValThrValSerGlyGlySerLeuSerAlaProHisGlyAsnV 338
2399 TTAGCCTGTGGCAATGCCCAAGCAACATTTAAT 2433
338 alIleGlyThrGlyGlyAlaArgArgPheProProProAlaSerPro 354
2434 ...CAAGCCACATTAAGCGCAACATCGGCTTCCGCAATGCT...TC 2477
355 LeuSerIleThrLeuGlnAlaGlyAlaArgAlaGlnGlyArgAlaLeuLe 371
2478 ATTTAATCTAAGCGACCGCTACAAAGGCAAGGCAAGTGTAGACGCTTTCG 2527
371 uTYrValValLeuProGlnProValTysLeuThrLeuAlaG 385
2528 GCACCGCTAAGCGCAACGTAAGCCATTCCGCACTCAACGTAATGTCCTC 2577
385 LyGlyAlaGlnGlyGlnGlyAspIleVal 394
2578 CTAGCCGATTAAGGACGACTTTCATTGTAAGCAGCCGCTTACGGACA 2627
395 AlaThrGluLeuProProIleProGlyAla 404
2628 AATCAGCGCGCAAGATACGCAATTAACATTAAGACACGCAATGA 2677
404 aSerSerGlyProLeuAspValAlaLeu...AlaSerGlnAlaArgTPT 420
2678 CGCTGCCCTCAGCAGCAGGAATTAGCAATTAAACCTTGACACCGCAC 2727
420 hr...GlyAlaThrArgAlaValAspSerLeuSerIleAspAsnAlaThr 435
2728 ATTACACTGAATTCGCCCTATCGCCACGATGGCGAGGCGCAACCGG 2777
436 TrpValMet 438
2778 CAGTGGACAGATGGCGCGCGCGCTTCGCGCGCTTCCGCGCTTCC 2827
439ThrAspAsnSerAsnValGlyA 446
2828 TATTATCGCTTACACCGCACTTGCTAGAA 2859
446 IAlaLeuArgLeuAlaSerAspGlySerValAspPheGlnGlnProAlaGly 462
2860 ...TCCCGTTCAACACGCTGACGTAAGCGCAAAATGAACGGTCAAG 2906
463 AlaGlyArgPheGlyValLeuMetValAsp...ThrLeuAlaGlySerGly 478
2907 AACATTCGCTTTATGTGGAATCTTTCGCTACGCAAGCAAAATGA 2956
478 yLeuPheArgMetAsnValPheAlaAspLeuGlyLeuSerAspLysLeuV 495
2957 AGCTGGCGGAAAGTTCGAGAGCACTTACACCTTCCGCGCTCAACAATAC 3006
495 alValMetArgAspAlaSerGlyGlnHisArgLeuTrpValArgAsnSer 511
3007 GCGCAAGCACTGCAAGCGCTGCAACAATTTGACGTAAGTGAAGAAAGA 3056
512 GlySerGlnProAlaSerGlyAsnThrMetLeuLeuValGln 525
3057 CAACAACCGCTGTCCGAAACCTTAATTTCACCTGCAAAACGA... 3102
526 ...ThrProArgGlySerAlaAlaThrPheThrLeuAlaAsnLysAspG 541
3103 ...CACGTGATGCGCGCGCTGCGGTACCAACTCATCCGCAAGAGCGG 3150
541 LyLysValAspIleGlyThrTyrArgTyrArgLeu 552
3151 GAGTTCGCGCTGCAATTCGCTCAAGAAACAGAGCTTCCGCAAACT 3200
552 552
3201 CGGCAAGGCAACCAAAAAACAGGGGGAAGAAAGACAGCGCAAGCC 3250
553AlaAlaAsnGlyAsnGlyGlnTrp. 560
3251 TTGACGGCTGATTTGCGCGCGCGCGATCCGCGCAAGAGCAAGAGC 3300
561 ...SerLeuValGlyAla...LysAlaProProAla...ProLysPr 573
3301 GTTCCGCAACCGCGCGCGAGCGAGCGGGGAAATGTCCGCAWTATGCA 3350
573 oAlaPro...GlnProGlyProGlnProGlyProGln 584
3351 GCGGAGAGAGAGAAAAACGGGTGCGAGGATTAAGACACGCGCTTG 3400
585ProPro... 586

```

3401 CGAAGACGCGGAGGAGGAAACCGCGCGCTTACACCGCGCTTCCCGCG 3450
587 .....GlnProGln..ProG 593
3451 GCGCGCGCGCGCGCGGATTTGCGCACTGACCAACCCCA..... 3492
593 InProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGln 609
3493 .CGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3541
609 aProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGln 625
3542 GTTGGAGGATTTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3585
625 LaValaThrGlyValGlyValGlyValGlyValGlyValGlyValGly 641
3586 CAGAGCGATTTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3627
642 SerSmaIaIeuSerIysArgIeuGlyGlyIleuArgIeuSnpProAsp 658
3628 ....CGCGTTTGGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3673
658 aGlyValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 675
3674 AAGATTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3723
675 rGAlaGlyArgArgPheAspIleValaIaIaGlyPheGlyIleuGly 691
692 AspHisAlaValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 708
3724 CAGAAAAACCTC.....GCGACGCGCGCGCGCGCGCGCGCGCG 3761
692 AspHisAlaValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 708
3762 TTCGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3808
708 aGlyThrThrArgIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 725
3809 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3858
725 hrAspSerValHisValaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 741
3859 TTTCACATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3906
742 PheTyrLeuAspAlaThrLeuArgIaIaIaIaIaIaIaIaIaIaIa 758
3907 .....TCAGACGCG.....ATCGGACGCGCGCGCGCGCGCGCG 3940
758 sValaIaGlySerAspGlyTyrAlaValaIaIaIaIaIaIaIaIaIa 775
3941 TCGTCGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3987
775 lYValGlyValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 791
3988 TTCGCGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4037
792 TrpPheLeuGlnProGlnProGlnProGlnProGlnProGlnProGln 808
4038 GGATTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4084
808 yLaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 825
4085 TCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4134
825 eValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 841
4135 CAACACATTTTCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4184
842 GlyGlyArgIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 858
4185 TTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4234
858 eAspGlyAlaGlyThrValaIaIaIaIaIaIaIaIaIaIaIaIaIa 875
4235 TCGGCAAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4284

```

```

..      |||||      |||
875 euArgIyThrArg...AlaGlyLeuGly..... 883
4285 TTCACGCTGTCCCTCCAGCTGCCCGCGCGCGCGCGCGCGCGCGCG 4334
884 .....LeuGlyMetAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 898
4335 G.....CAACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4371
898 aSerTyrGlyTyrSerIysGlyProIySleuAlaIaIaIaIaIaIa 912
seq_name: SwissProt_40:PERT_BORPE
seq_documentation_block:
ID PERT_BORPE STANDARD; PRT; 910 AA.
AC P14283;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Pertactin precursor (Outer membrane protein P.69) (P.93).
GN PRN OR OMP69A.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OX Bordetella.
OC NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-ISOLATE CN2992;
RA MEDLINE=89264462; PubMed=2542937;
RA Charles I.G., Dougan G., Pickard D., Chaffield S., Smith M.,
RA Novotny P., Morrissey P., Fairweather N.F.,
RA "Molecular cloning and characterization of protective outer membrane
RA protein P.69 from Bordetella pertussis."
RA Proc. Natl. Acad. Sci. U.S.A. 86:3554-3558(1989).
RL [2]
RP REVISIONS: TO 264 AND 332.
RX MEDLINE=92407514; PubMed=1527510;
RX Li J.L., Fairweather N.F., Novotny P., Dougan G., Charles I.G.;
RX "Cloning, nucleotide sequence and heterologous expression of the
RX protective outer-membrane protein P.68 pertactin from Bordetella
RX bronchiseptica."
RX J. Gen. Microbiol. 138:1697-1705(1992).
RL [3]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=96196517; PubMed=8609998;
RX Emsley P., Charles I.G., Fairweather N.F., Isaacs N.W.;
RX "Structure of Bordetella pertussis virulence factor P.69 pertactin."
RX Nature 381:90-92(1996).
CC -!- FUNCTION: AGGLUTINOGEN THAT BINDS TO EUKARYOTIC CELLS; A PROCESS
CC MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN
CC BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.
CC -!- SUBUNIT: MONOMER.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.
CC -!- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)
CC CONCENTRATION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: J04560; AAA22980.1; ALT_SEQ.
CC PIR: A32560; A32560.
CC InterPro: IPR003992; pertactin.
CC InterPro: IPR003991; pertactin_related.
CC PRINTS: PRO1482; PERTACTIN.
CC PRINTS: PRO1484; PERTACTINFAMILY.
CC Outer membrane; Signal; Virulence; Repeat.
CC SIGNAL 1 34

```


2680 CTGCCGTCAGGCAGCAATTTAGCAATTTAAACCTTGACAAAGCCACCAT 2729
::: ::::: |||||
425 ...GlyAlaThrArgAlaValAspSerLeuSerIleAspAsnAlaThrTr 440
2730 TACACTCAATTCGGCTTATCGCCAGCATGCGGAGGGCGCAACCGGCA 2779
::: :::::
440 pValMet..... 442
2780 GTGCACAGATGCGCGCGCGCTTCGCGCGCTTCGCGGTTCCCTA 2829
||||| ::::: |||||
443 ...ThAsp..... AsnSerAsnValGlyAla 450
2830 TTATCCGTTACACCGCCCACTTCGCTAGAA..... 2859
||| ::::: |||||
451 LeuArgLeuAlaSerAspGlySerValAspHegInGlnProAlaGluAl 467
2860 .TCCCGTTTCAACACGCTGACGGTTAAAGCGCAATTTGACGTCAGGAA 2908
::: ||||| ::::: |||||
467 aglyArgPheLysValLeuThrValAsn...ThrLeuAlaGlySerGlyL 483
2909 CATTCGCGTTTGTGCGAACTCTTCGCGCTACCGCAGCAACAATTTGAG 2958
||||| ::::: |||||
483 eupheArgMetAsnValPheAlaAspLeuGlyLeuSerAspLysLeuVal 499
2959 CTGCGCGAAAGTTCCGAAAGCACTTACACTTCGCGGTCACAAATATCCGG 3008
::: ::::: ||||| ::::: |||||
500 ValMetGlnAspAlaSerGlyGlnHisArgLeuThrValArgAsnSerGly 516
3009 CAGCAACCTGCAGCCTCGAACAATGACGTTAGTGCAGAAAGAAAGACA 3058
||::: ||||| ||||| ::::: |||||
516 LysSerIleProAlaSerAlaAsnThrLeuLeuValGln..... 529
3059 ACNAACCGCTGTCCGAAACCTTATTTACCTTCGCAAAACGAA..... 3102
||| ::::: |||||
530 ...ThrProLeuGlySerAlaIleThrPheThrLeuAlaAsnLysAspLys 545
3103 CACGTCGATCGCGCGCGCTTACCACTCATCCGCAAAAGACGGCGA 3152
||||| ||::: |||||
546 LysValAspIleGlyThrTyraArgLeu..... 556
3153 GTTCGCGCTGATATTCGCGGTCAAGAACAGAGCTTTCGACAAACTCG 3202
556 556
3203 GCNAAGCAGAAAGCCAAAGAGCGGAAAGACAGCGCAAGACCTT 3252
||| ::::: |||||
557AlaAlaAsnGlyAsnGlyGlnTrp... 564
3253 GACGCGCTGATTCGCGCGCGCGCGATCCGTCGAAAGACAGAAAGCGCT 3302
::: |||||
565 ...SerLeuValGlyAla.....Lys 570
3303 TSCCGAACCGCGCGCGGAGGAGCGGGAATGTCGCATTTATGACAG 3352
||| |||||
570 snAlaProProAla..... 574
3353 CGGAGGAAGAGAAAAAGCGGTGACGGGATAAAGACACCGCTTGCGG 3402
574 574
3403 AAACAGCGGAGGAGAAACCGCGCGCTACACCGCTTCCCGCGCGC 3452
::: ||||| ||| |||||
575ProLysProAla..ProGlnProGly..ProGln 584
3453 CCGCGCGCGCGCGGATTTGCCGCACTGCAACCCCAACCGGAGCGCC 3502
||||| ::::: |||||
585 ProProGlnProProGlnProGlnProGluAla..ProAlaProGlnPro 601
3503 AACCGCAGCGGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTAA 3552
||::: ||||| |||||
601 roAlaGlyArgGlnLeu...SerIleAlaIleAsnAlaValAlaSerThr 616
3553 TTTTCGCGCCAGCTCAACAGCGTT.....TTGCGCGTACAGGAATTT 3596

617 GlyGlyValGlyLeuAlaSerThrLeuTrpTyrAlaGlySerAsnAlaLe 633
::: ::::: ||||| ::::: |||||
3597 AGACCGGCTATTTCGCGAGACCGCGCCGCAAC.....GCCGTTT 3634
::: ||||| ||| |||||
633 LysArgLeuGlyGlyLeuArgLeuAsnProAspAlaGlyAlaTr 650
3635 GGACAAACCGGATCCGGGACACCAACACTACGCTTCGCAAGATTTCCGC 3684
||| ||||| ::::: |||||
650 rpeGlyArgGlyPheAlaGlnArgGlnGlnLeuAspAsnArgAlaGlyArg 666
3685 GCCTACCGCCCAACAAACCGACTGCGCCCAATCGTATCGCAAAAAACCT 3734
::: ||||| ::::: |||||
667 ArgPheAspGlnLysValAlaGlyPheGlyLeuGlyAlaAspHisAlaVala 683
3735 C.....GCCAGCGGCGGCTCGGCACTTCCTTTTCCCAACAC 3772
::: ||||| ::::: |||||
683 AlaValAlaGlyGlyArgTrpHisLeuGlyGlyLeuAlaGlyTrpThr 700
3773 GGACCGAAACACTTCGACGAGCGCATCGCAACTCGCACGCGCTTCC 3822
||| ::::: |||||
700 rgeGlyAspArgGlyPheThrGlyAspGlyGlyGlyHisThrAspSerVal 716
3823 CACGCGCGCGTTTTCGGCAATACGCCATCGACAG...TTCTACATCGG 3869
||| ::::: |||||
717 HisValGlyGlyTyrAlaThrTyrIleAlaAspSerGlyPheTyrLeuAs 733
3870 CATCAGCGCGCGCGGCTTTAGCAGCGGACCGCTT.....T 3907
::: |||||
733 pAlaThrLeuArgAlaSerArgLeuGlnAsnAspPheLysValAlaGlyS 750
3908 CAGACGCG.....ATCGAGGCAAAATCCCGCGCGCTGTGATTCAC 3951
||||| ::::: |||||
750 erAspGlyTyrAlaValLysGlyLysTyrArgThr.....His 762
3952 GGCATTACGACGATACCGCGCGCT.....TTGCGCGG 3986
||::: |||||
763 GlyValGlyLysLeuLeuGlyAlaGlyArgArgPheThrHisAlaAspI 779
3987 ATTGCGGATCGAACGACATCGCGCGCAACGCGCTTTCGCAAAAG 4036
::: |||||
779 TyrTrpLeuGlnProGlnAlaGlyLeuAlaValPheArgAlaGlyGlyG 796
4037 CGGATTAACGCTACGAAAC...GTCAATATCGCAACCGCGCTTCGA 4083
::: ||||| ||| ::::: |||||
796 LysAlaTyrArgAlaAlaAsnGlyLeuArgValArgAspGlyGlySer 812
4084 TTCACCGCTACCGCGCGGCGATTAAGCGAGATTTATTCATTCAAACGCG 4133
||| |||||
813 SerValLeuGlyArgLeuGlyLeuGlyValGlyLysArgIleGlyLeuAl 829
4134 GCNAACATTTTCATCAGCGCTTATTTGAGCGCTTCCTTACGATGCGG 4183
||| ::::: |||||
829 aglyGlyArgGlnValGlnProTyrIleLysAlaSerValLeuGlnLup 846
4184 CTTCGGGCAAAAGTCCGACACGCGGTCAATACCGCTATTTGGCTCAGAT 4233
::: ||||| ::::: |||||
846 heAspGlyAlaGlyThrValHisThrAsnGlyLysAlaHisArgThrGln 862
4234 TTGCGCAAAACCGCGAGTGGGAAATGGCGGTAAAGCCGAAATCAAGG 4283
::: ||||| |||||
863 LeuArgGlyThrArg...AlaGlyLeuGly..... 871
4284 TTTACGCGTTCCTCCAGCTGCGCGCGCAAGAGCCGCAACTGGAAG 4333
||||| ::::: |||||
872LeuGlyMetAlaIleAlaLeuGlyArgGlyHisSerLeuTyrAla 886
4334 CG.....CAACACAGCGCGGCGCATCAATTAGCTACCGCTGG 4371
||| ::::: ||||| |||||
886 LysTyrGlyTyrSerLysGlyProLysLeuAlaMetProTrp 900

seq_name: SwissProt_40:YFAL_ECOLI

seq_documentation_block:

ID YFAL_ECOLI STANDARD; PRT: 1250 AA.
AC P45508; P45507; P45506; P39441; P76468; P77487;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yfai precursor.
GN YFAL OR B2233.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubram S.,
RA Tagami H., Takahashi H., Takeda Y., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT -K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
RN [3]
RP PRELIMINARY SEQUENCE OF 1-938 FROM N.A.
RX MEDLINE=84272624; PubMed=6087316;
RA Carlsson J., Fuchs J.A., Messing J.;
RT "Primary structure of the Escherichia coli ribonucleoside diphosphate
RT reductase operon.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:4294-4297(1984).
RN [4]
RP SEQUENCE OF 925-1198 FROM N.A.
RC STRAIN-K12 / EMG2;
RX Estep P., O'Keefe T., Robison K., Church G.M.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1180-1250 FROM N.A.
RC STRAIN-OV6;
RX MEDLINE=88201664; PubMed=2834621;
RA Hussain K., Elliott E.J., Salmoud G.P.C.;
RT "The para-mutant of Escherichia coli also carries a gyraa mutation.
RT The complete sequence of gyraa.";
RL Mol. Microbiol. 1:259-273(1987).
RN [6]
RP IDENTIFICATION.
RX MEDLINE=96032851; PubMed=7567469;
RA Borodovsky M., McIninch J., Koonin E.V., Rudd K.E., Medigue C.,
RA Danchin A.;
RT "Detection of new genes in a bacterial genome using Markov models for
RT three gene classes.";
RL Nucleic Acids Res. 23:3554-3562(1995).
CC -!- SIMILARITY: TO E.COLI YDEK
CC -!- CAUTION: REF.3 SEQUENCE DIFFERS EXTENSIVELY FROM THAT SHOWN BY
CC MANY FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AE000313; AAC75293.1; -
CC EMBL: D90855; BAA16052.1; ALT_INIT.
CC EMBL: D90854; BAA16050.1; ALT_INIT.
CC EMBL: K02672; -; NOT_ANNOTATED_CDS.
CC EMBL: U00544; AA04094.1; -
CC EMBL: Y00544; -; NOT_ANNOTATED_CDS.
CC EMBL: E012850; yfai.
CC KMW Hypothetical protein: Repeat: Complete proteome.
CC FT SIGNAL 1 23
CC FT CHAIN 24 1250
CC FT DOMAIN 919 948
CC FT CONFLICT 28 30
CC FT CONFLICT 40 40
CC FT CONFLICT 65 66
CC FT CONFLICT 431 431
CC FT CONFLICT 433 434
CC FT CONFLICT 478 478
CC FT CONFLICT 773 773
CC FT CONFLICT 853 853
CC FT CONFLICT 923 924
CC FT CONFLICT 948 994
CC FT SEQUENCE 1250 AA; 131152 MW; 1798C05E299C95 CRC64;
CC
CC alignment_scores:
CC Quality: 282.00 Length: 1445
CC Ratio: 0.429 Gaps: 72
CC Percent Similarity: 45.467 Percent Identity: 20.208
CC
CC alignment_block:
CC US-09-303-518D-649 x YFAL_ECOLI ..
CC
CC Align seg 1/1 to: YFAL_ECOLI from: 1 to: 1250
CC
CC 637 AACCGGGAAGTTCATATCATATTCGAAGTCCGATTCT..... 675
CC
CC 46 SerArgGlnSerLeuSerGlyIleThrGlnAspIleAlaAspG1 62
CC
CC 676TGCGTGTGGTGGCAATACCTTGCACAAAGAGTGAATGAGTGGT 721
CC
CC 62 yGlnIrrpLeuValIlePheSerAsp...MetThrAsnAspAlaSerIyAla 78
CC
CC 722 GCACAGTCAACTAGTGTAGTGAATAAATTAACATAGCCCATATGCTTT 771
CC
CC 78 IalVal.....Phe 80
CC
CC 772 TTACCAACAGAGAGCGTCATT.....GCCGACAGTGGCTC 806
CC
CC 81 LeuGlnGlnIyAlaGluPheSerLeuLeuProGluAsnGluThrGlyMe 97
CC
CC 807 ACCAATGTTTATCTATGATGATGCCCAAAAGAGTGTAAATTAAGGGG 856
CC
CC 97 ThrLeuPheAlaAsnAsnThr.....ValThrGlyG 108
CC
CC 857 TATTGCAACAGCGCAACCCCTATATAGAAAGCAAGTGGCTTCCAGCTG 906
CC
CC 108 IuTyTrAsnAsnGlyIyAlaIlePheAlaIySgIuAsnSerThrLeuAsn 124
CC
CC 907 GTTGCTAAGATGTTGTTCTATGATGAATCTTGGTGGAGATCCCATTC 956
CC
CC 125 LeuThr.....AspValIlePheSerGlyAsnValAlaG1 136
CC
CC 957 AGTATTCTAGCAACACGTCGCAAAATGAGAAATCTTTTAAGCAGATA 1006
CC
CC 136 yGlyTyTr.....GlyGlyAlaIleTySerSerGlyT 147
CC
CC 1007 ATAATGCACAGCAAAATATCATATGCCAATGCATGCAAC..... 1044
CC
CC 147 hrAsnAspThrGlyAlaValAlaSpLeuArgValThrAsnAlaMetPheArg 163

705 TrpValGlyAspGlyAlaThrPheValThrGlyAlaAspGlnAspIleG 721
2547 AAGCATTCGGCACTC...AACGTAATGTCCTCCAGCCGATAGGACG 2593
721 nSerIleAspAlaIleSerSerGlyThrIleAspIleSerAspGlyThr 738
2594 TATTCATTTTGAAGACCGCTTACGGACAATTCAGCGCGCGCAG 2643
738 AlLeu.....ArgLeuThrGlyGln..... 744
2644 GATACGGCATTCACCTTAAGACACG.....GAATGACGCT 2681
745 AspThrSerValAlaLeuAsnIleSerLeuPheAsnGlyAspGlyThrIle 761
2682 GCGCGACGACCGAA.....TTAGCAATTTAAC..... 2712
761 ValAsnAlaThrAspGlyValThrIleuThrGlyLeuAsnThrAsnL 778
2713CTTGACACCGCACCATTCATCACTCAAT 2739
778 euGluThrAspSerLeuThrTyLeuSerAsnValThrValAsnGlyAsn 794
2740 TCCGCTATCGCCACGATCGCGACGGGCGCAACCGCAGTCGACAGA 2789
795 LeuThrAsnThrSerGlyAlaValSerLeuGlnAsnGlyValAlaGlyAs 811
2790 TCGCGCGCGCGCGCTTCGCGCGCTTCGCGCGCTTCCTATTATCCGTTA 2839
811 P..... 811
2840 CACCGCACTTCGGTAGATCCCTTTCACACCGCTAGCGTAACGCG 2889
812ThrLeuThrValAsnGly 817
2890 AATTGAAGGTCAGGAACATTCGCTTATGTCGAACTCTCGGC.. 2937
818 AspTYrThrGlyGlyThrLeuLeuAspSerGlyLeuAsnGlyAs 834
2938 ...TACCGCACGACAATTGAAGCTGCGGAAAGTTCCGAAAGCACTT 2983
834 pAspSerValSerAspGlnLeuValMetAsnGlyAsnThrAlaGlyAsn 851
2984 ACACCTTGGCGGTCAACAAT..ACCGGC...AACGAACCTGCAAGCTC 3027
851 hThrValValValAsnSerIleThrGlyIleGlyGlnProThrSerThr 867
3028 GAACAATTGACGCTAGTGGAAGAAAGACAACAACCGCTGCCGAAA 3077
868 GlyIleValValValAspPheAlaIleAspProThrGlnPheGlnAsnAs 884
3078 CCTTAATTCAACCTCGCAAAAGCAA...CACGTGATGCGCGCGCTGCG 3124
884 nAlaGlnPheSerLeuAlaGlySerGlyTYrValAsnMetGlyAlaTYrA 901
3125 GTTACCAATCATCCGCAAAAGACGCGAGTTCCGTCATTAATCCGCTC 3174
901 sPTyTYrThrLeuValGlnAspAsnAsnAspTYrLeu..... 913
3175 AAAGAACAGAGCTTTCGACAACTCGGACGAGCAAGCAAAAAACA 3224
913 913
3225 GCGGAAAAAAGACAAGCGCAACGCTTGAACGGCTGATGCGCGCGGCG 3274
913 913
3275 GCGATCCCGTCGAAAAGACAGAAAGCGTTGCGCAACGCGCGCGACAGA 3324
914ArgSerGlnGlnValThrProProser..... 922
3325 GCGCGGAAAAATGCGCATTTATGCAAGCGGAGAGAAAGAAAAAACGGGT 3374
922 922
3375 GCAAGCGGATAAAGACACACGCTTGGCGAAACAGCGCGGAGGAAACCC 3424
923 ProProAspProAsp..... 927
3425 GCGCGGCTACACCGCTTCCCGCGCGCGCGCGCGCGCGGATTTG 3474
928 ProThrProAsp..... 931
3475 CCGCAACTGCAACCCCAACCGGACCGCCACCGGACGACCTGATAG 3524
932 ProAspProThrProAspProAspProThrProAspProGlnProThrPr 948
3525 CCGTATGCGCAATAGCGGTTTGAATTTTCCGCGCAACCGCACACG 3574
948 oAlaTYrGlnProValLeuAsnAlaValGlyGlyTYrLeuAsnAsnL 965
3575 TTTTCCCGCTACAGACGATTAAGACCGGATTTTCCGCAACCGCGCC 3624
965 euArgAlaAla.....AsnGlnAlaPheMetMetGlnArgTYr 977
3625 AACCGCGTTTGGACAAAGCGC.....ATCCGGGA 3653
978 AspHisAlaGlyGlyAspGlyGlnThrLeuAsnLeuArgValIleGlyG 994
3654 CACCAACACATACCGTTCCGAAGATTCCGCGCTACCGCAACAAACG 3703
994 yAspTYrHisTYrThrAlaAlaGlyGlnLeuAla.....GlnHisGln 1009
3704 ACCTGCGCAATCGGTATGCGAAGAAACCTCGCAGCGCGCGC..... 3747
1009 sPTThrSerThrValGlnLeuSerGlyAspLeuPheSerGlyArgTYrGly 1025
3748GTGCGCATCCTG.....TTTGCGCAACCG 3773
1026 ThrAspGlyGlnTYrPheLeuGlyIleValGlyGlyTYrSerAspAsnG 1042
3774 GACGCAAAACACTTCGACGACGCGCATCGGCACTCGGACGCGTTGCC 3823
1042 nGlyAspSerArgSerAsnMetThrGlyThrArgAlaAspAsnGln 1059
3824 ACGGCGCGCTTTCGCG.....CAATCGGATCGAC 3855
1059 IsGlyTYrAlaValGlyLeuThrSerSerThrPheGlnHisGlyAsnGln 1075
3856 AGGTTCTACATCGGATCAGCGCGCGCGGCTTTAGC.....AGCG 3899
1076 LysGlnGlyAlaTYrPheAspSerTYrPheGlnTYrAlaThrPheSerAs 1092
3900 CAGCTTTCAGACGCGATCGGAGCAAAATCCCGCGCGCTGCTGCAT 3949
1092 nAspValSerGlnGlnGlnAspGly.....ThrAspHis 1104
3950 AC.....GGCATTCAGGACGATACCGCGCGCGCTTC 3981
1104 yHisSerSerGlyIleAlaSerLeuGlnAlaGlyTYrGlnTYrPhe 1120
3982GGCGATTCGCGATCGAACCGCACATCGGCAACGCGTATTT 4025
1121 ProGlyArgGlyValValIleGlnProGlnAlaGlnValIleTYrGlnG 1137
4026 CGTCAAAAAAGCGATTCAGCTACGAAAC..... 4056
1137 yValGlnGlnAspAspPheThrAlaAlaAsnAlaGlyAlaTYrSerGln 1154
4057GTCATATGCGCAACCGCGCGCTTCATTCACCGTACCGC 4098
1154 euGlnGlyAspAspIleGlnThr.....Arg 1162
4099 GCGGCATTAAGGAGATTTATTCATCAACCGCGCGCAACATTTCCAT 4148
1163 LeuGlyLeuHisSerGlnTYr.....ArgThrAlaValHis.....Va 1175


```

310  ....: |||
    ertgltthrglnasnlleryserglyserlaleuserlalaasnille 326
1390 AAGGACCGTGCACGTTCAAGCCAAAGGGAA..... 1422
    |||||: |||: |||
327 lsglysercglntlevalasnserglnglythrallaleasnthrleuva 343
1423 ..... AACCAAGGCTGCATCAGGCGGGG 1447
    :|||: |||: |||
343 lseraspglytyrtyrlnhstlleargsnnglyllealasergly. 359
1448 ACGGTACGATCTTTGGATCAGACGACGACGATTAAGCAAAAAACAA 1497
    :|||: |||: |||
360 ..... Thrilevalasnnglnserglytyrvalasnillesergly 373
1498 GCCTTAGTGAATCGCTTGTCACGCGGACGCGGTGACGTGCACACTGA 1547
    |||||: |||||: |||||
374 glytyrlnhstlserthrilleasnserglytyrthrleuargvalle 390
1548 TCGCGATTAATCAGTTCACCCGACAACTTATTCGGCTTCGGGCGG 1597
    :|||: |||: |||
390 useraspglytyrtyrlnhstlserglytyrthrleuasn..... AsnSerg 404
1598 GACGTTGATTTAAAGCGGACATTCGCTTCACCGATTCGAAAT 1647
    |||||: |||||: |||||
404 lylarglnasnvalserasnnglylyvalserlyrasmalmetileasn 420
1648 ACCGATGAAGGCGCGATGTCATTCACCAACCAAGCAAGCAAGATCCAC 1697
    |||||: |||||: |||||
421 Thrnglylyasnnglnlyrtyrtyrseraspglylnalathralalal 437
1698 CGTTACCATTTACAGCAATTAAGATATTCCTCAACCGGCAT..... 1740
    |||||: |||||: |||||
437 evalasnthrserglypnehlarglleasnserglytyrthrleuprov 454
1741 .. AACACAGCTTGATGACCAAAAAAAGAAATGCC..... 1773
    :|||: |||: |||
454 alglinsnservalvalthrargthrvalseralalalyspro 470
1774 ..... TACAAACGCT..... 1782
    |||||: |||||: |||||
471 Pheaspalaglvaltyrserglytyrlysglnthrvaltyrleutprar 487
1783 ..... TGCTTGGCGAGAAAGATACG..... 1803
    |||||: |||||: |||||
487 gcllyltyrtyrserasnphelutthrvalaltrpsermetpneprog 504
1804 ..... ACCAAAGCAAGCGGCGGCTCAAC..... 1827
    :|||: |||: |||
504 lylthralserglylalaasnvalasnleuserglyltyrleuasnlaphe 520
1828 ..... CTTGT 1832
    |||||: |||||: |||||
521 Alaglyasnvalvalglythrilleuasnnglnnglyltyrarglntyva 537
1833 TTACAGCGCGCGCGCA..... GAGACCGCA 1858
    |||||: |||||: |||||
537 llysercglalathralathrserthvalglyasnasnnglnnglyltyr 554
1859 CCCTGCTCTTCCGCGGCAACAAATTTAAAGCAACATCAGCAACAA 1908
    :|||: |||: |||
554 lutyvalleusercglylyltyrthr... Aspglythrvalleuasnser 569
1909 AAGCGCAACATGTTTTCAGC..... GCGACCAACACCGCACGCCCTA 1952
    :|||: |||: |||
570 glytyrleuglnalavalsercglytyrlyslaserlathrvalil 586
1953 CAATCATTTAAAGCAACATTTGTGCAAAAAGAGGCAATTCCTCGCGGG 2002
    |||||: |||||: |||||
586 easn..... Glyly... 592
2003 AATTCGTGTGGACAAACGATGATCAACCGACATTTAAAGCGGAAAC 2052
    :|||: |||: |||

```

```

592 lnphevaltyrtyrlytyrlyvalthrlythrlnstlleyasnngly 608
2053 TTCCAAATTTAAAGCGGACAGCGGTGTTCCGCAATGTTGCC... AA 2099
    |||||: |||||: |||||
609 glythrilleargvalaspserglylaserlaleuasnillealaleuse 625
2100 AGTGAAGCGCATTTGGCATTTGACCAATCAGCC..... 2133
    :|||: |||: |||
625 iserglylyasnleuphenthserthrlyalathrleuproglutaur 642
2134 ..... CAAGCAGTTTGTGTGCGACCG... CATCAAGCGCACACATC 2175
    :|||: |||: |||
642 hrtthmetlaleuaservalserglinaasnhtlaserlaleval 658
2176 TGTACACGTTGCGACTGACGCGGCTGTGCAAAATGTGTGCAAAAACAT 2225
    :|||: |||: |||
659 leuglnasnnglylyrleuargvalthrser..... Glylythr 673
2226 TACGACGATTAAGTATGCTTCATTCATGACTTAACACGACATTCAGCGCA 2275
    |||||: |||||: |||||
673 athrasp..... Thrthrvalasnserlaleyla 683
2276 ATGTGATCTTGCAGATCAGCTCATTTAAATTCACAGGCGCTTGCCACA 2325
    :|||: |||: |||
683 rgleuarglleaspaspglytyrthrileasn..... Glythrthrthr 697
2326 CTCACGCGC..... 2334
    :|||: |||: |||
698 lleasnlaaspglylylevalalaglythrlnstlleglnasnaspglyas 714
2335 ..... AATCTAGTGAATGCGCATTCACGCTTATACAGTCAAGCC 2374
    |||||: |||||: |||||
714 nphelleuasnleuvalaglnsentrlyrpneghlythrleuasnserg 721
2375 ACAAGCC..... ACCCAAAAGCGC 2394
    :|||: |||: |||
731 lyserglyvalleuvallylaspasnthrlylmetethrtyrlyalagly 747
2395 AACCTTAGCCCTC..... GTGCGCATGCCCCAACCAACAT 2429
    :|||: |||: |||
748 Thrleuthrglnalaglnnglyvalasnvallyllyllyltyrleuph 764
2430 TTAATCAAGCCCAATTAAGCGCAACACATCGGCTTGCGCAATGCTTCAT 2479
    :|||: |||: |||
764 easpserlalevalvalasnlaaspmetlalevalasnnglnalalatyrt 781
2480 TTAATCTAAGCGACCGCGCTACAAAGCGCAGTCTGACGCTTTCGCGC 2529
    :|||: |||: |||
781 leasnilleaseraspglinalathrilleasnnglyser..... Valasn 795
2530 AAGCTAAGCAAAACGTAAGCCATTCGCGCACTGCAAGGTAATGTCCCT 2579
    :|||: |||: |||
796 Asnnglyserllevalilleasnasnserlletlleasnnglyllyltyrthr 811
2580 AGCGATTAAGCAGATTTCCATTTTGAAGCAGCGCTTACCGGACAAA 2629
    :|||: |||: |||
812 ..... Asnaspalaspserpneghlythrlylals..... 822
2630 TCAGCGCGCGCAAGATACGCGCTTACACTTAAAGACACGCAATGACG 2679
    :|||: |||: |||
822 ..... 822
2680 CTGCGGTACAGCAGCAAGATTAAGCAATTTAAACCTTGACACGCGCCAT 2729
    :|||: |||: |||
823 leuasnserlathrvalasnnglyserleu... Valasnlnlyllyltyrthr 838
2730 TACACTCAATTTCCGCTATTCGACAGATGCGGACGCGGCAAAACCGGCA 2779
    :|||: |||: |||
838 elileuasn..... 841
2780 GTGCGACAGATGCGCGCGCGCGCGCTTCCGCGCTTCCGCGCTTCCCTA 2829
    :|||: |||: |||
841 ..... 841

```

```

2830 TTATCCGTACACGCCCACTGGTAGAATCCCGTTTCAACAGCTGAC 2879
      |||||::: ||||| |||||
842 .....ProthrLys...GluSerAlaGlyAsnThrLeuTh 852
2880 GGTAAACGGCAAAATGTAACGGCTGAGGAGACATTCGCTTATATGCGAAC 2929
      |||||::: ||||| |||||
852 rValSer.....AsnTyrThrGlyThrProGlySerValLleSerL 866
2930 TCTTCGGCTAC.....CGCAGCGACAATTAAGACTG 2961
      |||||::: ||||| |||||
866 euGlyGlyValLeuGluGlyAspAsnSerLeuThrAspArgLeuValAl 882
2962 GCGGAAATTCGCAAGGCACTTACACCTTGGCG...GTCAACATATCCG 3008
      |||||::: ||||| |||||
883 LysGlyAsnThrSerGlyInSerAspLleValTyrValAsnGluAspGl 899
3009 CACGACCTGCAAGCCTCGAACATTAAGCGTAGTGGAAGGAAGACA 3058
      |||||::: ||||| |||||
899 ySerGlyGlyInThrArgAspGlyLleAsnLleSerValGluGlyA 916
3059 ACAAAACCGCTGCGCAAAACCTTAATTCACCTGCAAAACGACACGTC 3108
      |||||::: ||||| |||||
916 sn.....SerAspAlaGluPheSerLeuLysAsn...ArgVal 927
3109 GATGCGGCGGCGGTGCGCTTACCACTCATCCGCAAGAGCGGAGATTCCG 3158
      |||||::: ||||| |||||
928 ValAlaGlyAlaTyrAspTyrThrLeu...GlnYsgLysAsnGluSerGl 943
3159 CCGTCATATCCGCTCAAAAGAACAGACTTTCGCAAACTCGCGCAAG 3208
      |||||::: ||||| |||||
943 yThrAspAsn.....LysGlyTyrPtyrLeuThrSerHisLeuProThr 958
3209 CAGAAGCCCAAAAACAGCGGCAAAAGACAGCGCAAGCCTGACCGC 3258
      |||||::: ||||| |||||
958 eAspThrArgGlnTyrArgProGluAsnGlySerTyrAlaThrAsnMet 974
3259 CTGATTGCGGCGGCGGCGATGCCGTAAGAAAGACAGAAAGCTTGCGCA 3308
      |||||::: ||||| |||||
975 AlaLeuAla..... 977
3309 ACCGCGCGGCGAGCGAGCGGGAATGTGCGCATTAATGCAAGCGGAGG 3358
      |||||::: ||||| |||||
978 .....AsnSerLeuPheLeuMetAspLeuA 986
3359 AAGAAAAAACGGGTGCGAGCGGATAAGACAGCCGCTTGCGCAACAG 3408
      |||||::: ||||| |||||
986 snGluArgLysGlnPheArgAlaMetSerAsp..... 996
3409 CGCGAAGCGGAACCGCGCGGCTACCAACGCGCTTCCCGCGCGCGCG 3458
      |||||::: ||||| |||||
997 .....AsnThrGlnProGluSerAla..... 1003
3459 CGCGCGCGCGGATTTGCCGAACCTGCAACCCCAACGCGACCCCAACGC 3508
      |||||::: ||||| |||||
1003 ..... 1003
3509 AGCGGACCTGATCAGCCGTATATGCAATAGCGGTTAGTGAATTTTC 3558
      |||||::: ||||| |||||
1004 .....SerValTrpMetLysLleThrGlyGlyLies 1014
3650 GGGAGACCAAAACACTACGCTTGCAAGATTTCGCGCTACCGCCACAA 3699
      |||||::: ||||| |||||
1014 eSerGlyLys..... 1017

```

```

3700 ACCGACTCGCGCAAAATCGTAGTACGAAAAAACTCGCGAGCGGCGCT 3749
      |||||::: ||||| |||||
1018 .....LeuAsnAspGlyLyn... 1022
3750 CGGCATCTGTTTTCGACACACCGGCAAAACACTTTCGACAGCGCA 3799
      |||||::: ||||| |||||
1023 .....AsnLysThrThrAsnGlnPheLleAsnGlnL 1034
3800 TCGGCACATCGGAGCGGCTTGCCACAGCGCGCTTTCGGGCATATACGC 3849
      |||||::: ||||| |||||
1034 euGlyValAspLleTyrLysPheHisAlaGluInLeuGlyAsp..... 1048
3850 ATCGACAGGTTCATACATCGCATACAGCGGCGCGGCTTATGACAGCG 3899
      |||||::: ||||| |||||
1049 .....PheThrLeuGlyLleMetGlyGlyTyrAlaAsnAlaLysGl 1062
3900 CACCTTTACAGCGGATCGAGGCAAAATCCGCGCGCGCTGCTGCAT 3949
      |||||::: ||||| |||||
1062 yLysThrLleAsnTyrThrSerAsnLysAlaAlaArgAsnThrLeuAsp 1078
3950 ACGCATTCAGGACGATACCGCGCGGCTTTCGGCGGATTCGCAATGCA 3999
      |||||::: ||||| |||||
1079 .....GlyTyrSerValGlyValTyrGlyThrTrpTyrGln 1090
4000 CCGCACATCGCGCAACCGGCTATTTCGTCAAAA...CGCGATTACCG 4046
      |||||::: ||||| |||||
1091 AsnGlyGluAsnAlaThrGlyLeuPheAlaGluThrTrpMetGlnTyrAs 1107
4047 CTACGAAACGTCAATATCGCCAGCCCGCGCTGATTCACACGGTACC 4096
      |||||::: ||||| |||||
1107 nTrpPheAsnAlaSerValLysGlyAspGlyLeuGluGluLysTyrA 1124
4097 GC..... 4098
1124 snLeuAsnGlyLeuThrAlaSerAlaGlyGlyTyrAsnLeuAsnVal 1140
4099 .....GCGGCATTAAGCGCAGATTAATTCATCAACAC 4130
1141 HisThrTrpThrSerProGluGlyLleThrGlyGluPheTrpLeuInPr 1157
4131 G.....GCGCACACATTCATCCATCAGCGCTTATGAGCTGTCT 4171
      |||||::: ||||| |||||
1157 OHsLeuGlnAlaValTrpMetGlyAlaInProAspThrHis..... 1171
4172 ATACCGATGCGGCTTCGCGCAAGTCCGAACCGCTCAATACGCGCTA 4221
      |||||::: ||||| |||||
1172 .....GlnGluAspAsnGlyThr 1177
4222 TTGGCTACGATTTTCGCAAAACCGCG.....AG 4250
      |||||::: ||||| |||||
1178 ValValGlnGlnAlaGlyLysAsnAsnLleGlnThrLysAlaGlyLleAr 1194
4251 TCGCGAATGCGGCTTAACCGCGCAAAATCAAAAGTTTACGCTGCTCC 4300
      |||||::: ||||| |||||
1194 gAlaSerTrpLysValLysSer.....ThrLeuAspLysA 1206
4301 ACCTGCGCGCGCGCAAGCGCGCAACTGGAACGCA..... 4338
      |||||::: ||||| |||||
1206 sPthrGlyArgArgPheArgProTyrLleGlnAlaAsnTrpLleHisAsn 1222
4339 ...CACAGCGCGGCGCATCAATTA 4359
      |||||::: ||||| |||||
1223 ThrHisGluPheGlyValLysMet 1230

```

seq_name: SwissProt_40:YEJO_ECOLI

seq_documentation_block:
 ID YEJO_ECOLI STANDARD: PRT; 863 AA.
 AC P33924; P76450; Q47291;
 DT 01-FEB-1994 (Rel. 28; Created)
 DT 01-FEB-1994 (Rel. 28; Last sequence update)
 DT 16-OCT-2001 (Rel. 40; Last annotation update)
 DE Hypothetical outer membrane protein yeJo.

GN YEJO OR B2190.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision: Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_Taxid=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / BHB2600;
 RA Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,
 RA Church G.M.;
 RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; Pubmed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP PRESENCE OF AN INSERTION SEQUENCE.
 RA Rudd K.E.;
 RL Unpublished observations (JAN-1994).
 CC -1- SUBCELLULAR LOCATION: Outer membrane (Potential).
 CC -1- SIMILARITY: STRONG, TO BORDETELLA PERTACTIN.
 CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE READING FRAME IS
 CC INTERRUPTED BETWEEN CODONS 21 AND 22 BY A 155K INSERTION ELEMENT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U00008: AAL16385.1: ALT-SEQ.
 DR EMBL: AE000308: ACCT5250.1: ALT-SEQ.
 DR Ecogene; Egi2051: yejo.
 KW Hypothetical protein; Outer membrane; Complete proteome.
 SQ SEQUENCE 863 AA; 91202 MW; 55D4600B5BC3D94D CRC64;

alignment_scores: Quality: 265.50 Length: 1023
 Ratio: 0.575 Gaps: 49
 Percent Similarity: 45.161 Percent Identity: 20.723

alignment_block:

US-09-303-518D-649 x YEJO_ECOLI ..

Align seg 1/1 to: YEJO_ECOLI from: 1 to: 863

```

1693 TCACACCGTT.....ACCATACAGC.....ATAAAGA 1721
|||||
30 Serthrvalglutryglutthivalasplvalvalleuileuilsas 46
1722 TATTGCTACA.....ACGGCATAAACAAGCTTGATAGCAAAA 1762
|||||
46 pilleglleuvaltyrlyglthralasnasnthrllysileasproglyg 63
1763 AAGAAATGGCTACACAGGTTGGTGGAGCAAAAGATACGACCAAAACG 1812
|||||
63 lyglu...gluhstlelysgluhpheglyvalserasnasnthrgluile 78
1813 AAGCGCGGCTCAACCTTTGTTACAGCCGCGGAGAGACCGACCGACCT 1862
|||||
79 Asmolyltyrlyglintyrlieglumetasnglyalalaglutrysera 95
1863 GCTGCTTCGCGCGGA.....ACAATTAACGCAACATCAGCAAA 1906
|||||

```

```

95 lleuasnaspglytyrlygllevalglumetglylyalalaaasngint 112
1907 CA.....AAGCGCAAACTGTTTTCAGCGGAGACCAACACCGAC 1947
|||||
112 hrthrleuasnasnglyvalleu.....glu 120
1948 GCCTACATTCATTTAAAGCACCATTGGTCGCAAAAAGAGGCAATTCCTG 1997
|||||
121 valtyrlyalalalaaasnasp.....Thrthrilleysgl 132
1998 CGGCGAAATCGTGTGGACACACTGATACCGCATTTAAAGCG 2047
|||||
132 yeltyrlyglleuilevalglulysaspgly...glyalavalphevalalai 148
2048 AAAACTTCCAAATTTAAAGCGGAGCGGCGGTGGTTCCCGCAAT..... 2091
|||
148 lglu.....lysglyglyleuileuvallysglyglglyphe 161
2092 .....GTGCCAAAGTGAAGCGCATTTGGACCAATCA 2129
|||||
162 Alaiphealavalaspglnlysalaglyalalaleysthrthrtr 178
2130 CGCCCAAGCAGTTTGTGTGTC..... 2151
|||||
178 galmetgluvalpheglythrthasnarlygluylpheaspilleysa 195
2152 ..... 2154
195 snglytlealaaasnasmetleuengluasnglylyserleuargval 211
2155 CGGCATCAAGCGCACACATCTGACGCTGGAGCTGAGCGGTCTGAC 2204
|||||
212 glugluasnasphthalatyrasnthrthvalaspsersglyleuile 228
2205 AAATGTCTCGAA...AAACCATTACC.....GACG 2233
|||
228 ugluvalmetaspglyglythrvalthnglyvalasplylysalaaglyg 245
2234 ATAAAGTATGCTTCATTCATTCATGACCAAGCAGCATACCGCAATGTCAT 2283
|||||
245 lylysleuilevalserthrthasnalaleuenglualserglypro..... 259
2284 CTGGCCGATCAGCGCTCAATTTAATCTCAGAGCGCTGCCACACTCAACGG 2333
|||||
260 .....Asnserlysglylpheserilleysas 269
2334 CAATCTTAGTGCATAT.....GGCGATACAGCTTATACAGTCAGCC 2374
|||||
269 pelyvalserlysasntrygluleuasnaspglyserglyleuilevalm 286
2375 ACAAGCGCACCCCAAAACGGCACTTGCTGCGCAATGCCCAAGCA 2424
|||
286 etgluaspthngln.....Alalleasphthrilleuasplyshsala 300
2425 ACATTTAATCAAGCCACATTTAAACGGCAACATCGGCTGGGCAATGC 2474
|||||
301 Thmetglnserleuenglylysaspthnglythrlyvalglalalasnml 317
2475 TTCAATTAATCAAGCAGCACCACCGCTGACAAAGCGCACTGACGCTTT 2524
|||||
317 aavaltyraspleuengly...argserlyrghnasnglyserillethrtyrs 333
2525 CC.....GGCAACGCTTAAGCTTAAC 2544
|||
333 erserlysalalalesergluasmetvalilleasnasnglyaralasn 349
2545 GTA.....AGCCATTCCGCACTCAACGGAATAGTCTCCCT 2579
|||||
350 Valtyrpalaglythmetvalasnalservalarvglylasnaspglyyl 366
2580 AGCGCAAT.....AAGCAGTATTCATTTTGAAGAAGCCGCTTACCG 2623
|||
366 elaugluvalmetlysproglilleasnthryalaproalmetleuvalg 383

```



```

|||||
817 laglythrseryValtyrGlnAlaGlyLeaArgSerSerPheThrPro 833
4288 ACGGTGTCCTCCACGCTGCGCGCGCAAGCGCCCGCACTGACGCA 4337
|||||
834 ThrleuSerGlyHisLeuSerValSerTyrGly..... 844
4338 ACACAGCGCGGCATCAAA 4356
|||||
845 .AsnGlyAlaGlyValGln 850

seq_name: SwissProt_40:YDBA_ECOLI

seq_documentation_block:
ID YDBA_ECOLI STANDARD; PRI: 2003 AA.
AC P33666; P76087; P76088; P76856; P76857; P76859;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ydbA.
GN YDBA OR B1401/B1405.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sempel G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,
RA Takeuchi K., Takeuchi Y., Wada C., Yamamoto Y., Horiiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [3]
RP SEQUENCE OF 464-2003 FROM N.A.
RC STRAIN-K12;
RX MEDLINE=92190338; PubMed=1665988;
RA Moszer I., Glaser P., Danchin A.;
RT "Multiple IS insertion sequences near the replication terminus in
Escherichia coli K-12.";
RL Biochimie 73:1361-1374(1991).
CC -!- SIMILARITY: TO S. TYPHIMURUM ORF NEAR CYS6 (AC P25928).
CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE GENE CODING FOR
THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT
BETWEEN AMINO ACIDS 839 AND 840.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000237; AAC74483.1; ALT_SEQ.
DR EMBL: AE000237; AAC74487.1; ALT_SEQ.
DR EMBL: D90778; BAA15009.1; ALT_SEQ.
DR EMBL: D90778; BAA18880.1; ALT_SEQ.

```

```

DR EMBL: D90779; BAA18881.1; ALT_SEQ.
DR EMBL: X62680; -; NOT_ANNOTATED_CDS.
DR Ecogene; EG1307; ydbA.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 489 489 I -> V (IN REF. 2).
FT CONFLICT 495 495 I -> V (IN REF. 2).
SQ SEQUENCE 2003 AA; 205949 MW; B83A12C8B53220FE CRC64;

alignment_scores:
Quality: 252.00 Length: 1864
Ratio: 0.311 Gaps: 89
Percent Similarity: 43.401 Percent Identity: 19.152

alignment_block:
US-09-303-518D-649 x YDBA_ECOLI
Align seg 1/1 to: YDBA_ECOLI from: 1 to: 2003

172 GCCGAATAATTAAGCAAGTTTGCAGTCGGGCGCAAGAT.....AT 212
|||||
324 AlaAsnAsnAsnGlySerThrValAspGlyLysAspSerThrClyTh 340
213 TGAGTTTACACAAAAGGAGGTGTC..... 243
340 rGluIleAsnGlyAsnAsnGlyValIleGlnAspGlyLysLeuAsp 357
244 .....GGCAATTCATGACA... 258
357 alserGlyGlyHisGlyIleAspIleThrGlyAspSerAlaThrVal 373
259 .....AAGCCCCGATGATT..... 273
374 AspAsnLysGlyThrMetThrValThrAspProGlnSerIleGlyIle 390
274 .....GATTTTCGTGTCGGGCAAGCCGTCGGCATTCG 313
390 nValAspGlyAspGlnAlaValAlaValAsnAsnGlyLysIleAlaIle 407
314 TG.....GGCATCATATATGTGAGC 336
407 hrAsnGlyGlyThrGlyThrGlnIleAsnGlyAspAsp.....Ala 420
337 GTGGCAATTAAGCGGCTATTAACAGCTTATTTGTGCGGAGAGAG 386
421 ThrAlaAsnAsnGlyLysThrThrValAsp...GlyLysAspSerTh 436
387 AATCCGATCAACATCGTTTACTTATTAATTAATTAATTAATTAAT 436
436 rGlyThrGluIleAlaGlyAsnAsnGlyLysValIleGlnAspGlyAsp 453
437 ATAAAGCGAGGACTTAAGCCAT.....CCTATGCGCGCATTAATCAT 480
453 euAspValSerGlyGlyHisGlyIleAspIleThrGlyAspSerAla 469
481 ATGCGCGCTTGCAATTAATTAATTAATTAATTAATTAATTAATTA 530
470 ThrValAspAsnLysGlyThrMetThrValThrAspProGlnSerIle 486
531 CAGTTATATGATGAGCGGCAATATATATCAATTAATTAATTAATTA 580
486 yIleGlnIleAspLysGlnAlaIleValAlaAsnAsnGlyLysIle 503
581 GTGTT...CGTATGGGCGAGCGGCAATATTTGGCATGTGTAAGAT 627
503 hrIleThrAsnGlyThrGlyThrGlnIleAsnGlyLysAspAlaThr 519
628 GAGCCCAATAACCGCAAGTATCATATCATATGCAAGTGCATCTCTG 677
520 AlaAsnAsnSerGlyLysThrThrValAspGlyLysAspSerThrCly 536
678 GCTGTGTGGTGCATACCTTTGCACAAAATGATGATGATGATGATG 727
|||||

```

536 rLysIleAlaGlyAsnIleGlyIleValAsnLeuAspGlySerLeuThrV 553
728 TCACATTAGTACTGAAAAATTAACATAGCCCATATGCTTTTACCA 777
553 alThrGlyGlyAlaHisGlyValGlu..... 561
778 ACAGAGGCTCATTTGGCGACAGTGGCTCA.....CCAA 812
562AsnIleGlyAspAsnGlyThrValAsnAsnGlyAspIle 575
813 GTTATCTATGATGCCCAAAAGCAAGTGTATTAATGAG..... 855
575 evalValSerAspThrIleGlyValLeuIleAsnGlyGluGlyA 592
856GTATTCAAACGGCGACCCCTATATA..... 882
592 laThrValSerAsnThrGlyAspValAsnValSerAsnGluAlaThrGly 608
882 882
609 PheSerIleThrThrAsnSerGlyLysValSerLeuAlaGlySerMetG 625
883GGAAGCAATGGCT 898
625 nValGlyAspPheSerThrGlyValAspLeuAsnGlyAsnAsnAsnSerV 642
899 TCCAGCTGGTTCGTAAGATTGGTTCTATGATGAATC..... 936
642 alThrLeuAlaIleLysAspLeu.....LysValValGlyGlnLys 655
937TTTGTGGAGATACCCATTCAGTA..... 960
656 AlaThrGlyIleAsnValSerGlyAspAlaAsnThrValAsnIleThrG 672
961T 961
672 yAsnValLeuValAspLysAspLysThrAlaAspAsnAlaIleGluTyrP 689
962 TCTACGAAACACGTCMAATGGAAATATCTTTTAACGACGATATAT 1011
689 hepPheAspPro...SerValGlyIleAsnValTyrGlySerAspAsnAsn 704
1012 GGCACA.....GGAATAATCATGCCCCAATGACACAAATTCCTGGCC 1055
705 ValThrLeuAspGlyLysLeuThrValValSerAsp..... 716
1056 TAATGATTAATAACACGACCGCTCAATGTTATGTTCTTCTTATCCG 1105
717 .SerGluValThrSerArgGlnSerAsnLeuPheAspGlySerAlaGluL 733
1106 AGACAGCAAGACACCTGTTTATCATGCTGCGAGTGGTGTCAAC..... 1149
733 ySerThrSerGlyLeuValValIleGlyAspGlyAsnThrValAsnMetAsn 749
1150AGTTATCGACCCAGCTGATTAATGAGAAAA 1181
750 GlyGlyLeuGluLeuIleGlyGluLysAsnAlaLeuAlaAspGlySerG 766
1182 TATTTCTTATTTGACGAAAGAAAGC.....GAATTGATCTTACCA 1225
766 nValThrSerLeuArgThrGlyTyrSerThrSerValIleValValS 783
1226 GCAACATCAATCAAGTGGTGGAGATTTATTTCCAGAGATTTTACG 1275
783 er.....GlyGluSerSerValTyrLeuAsnGlyAspThrThr 795
1276 GTCCTGCCCTGAAAAATTAACGAACTGGCAAGC..... 1308
796 IleSerGlyGluPheProLeuGlyPheAlaGlyValIleArgValGlnAs 812
1309GGGGCGTTTCATATCAGTGAAGACGATACCGTTACTTGGAAAGTAA 1354
812 pLysAlaLeuLeuGluIleGlySerGlyValaThrLeuThrMetGln.... 827

1355 ACGGCTGGCAACAGACCGCCTGTCCAAATCGCCAAAGCAGCCTGCAC 1404
828AspIleAspSerPheGlyIleHisGlyThrArgThrValGlu 841
1405 GTTCA.....GCCAAAGGGGAAAAACCAAGCTCGAT 1436
842 IleGlnAsnLeuGlyPheAlaPheValThrGlyGluAsn.....ThrTh 856
1437 CAGCTGGCGCAGCTACAGTCATTTGGATCAGCAGCAGCAGATTAAG 1486
856 rGlyIleAsnSerGlyThrIleSerLeuGln.....AsnG 869
1487 GCAAAAAACAAGCCTTTAGTGAATC..... 1512
869 LyLysAspProAlaProSerProIleValIleLeuAlaThrAsnGlyGly 885
1513GGCTGGTCAAGCGCGACGGGTACGGTGCACATCA 1547
886 SerAlaThrAsnAlaGlyThrIleThrGlyLysValThrGluGlnHisE 902
1548 TGGCGATATACGTTCAACCCC...GACAACTCATTTCCGCTTCCGCG 1594
902 rValPheAsnLysTyrSerThrGlyThrSerAsnSerPheIlePheAsnA 919
1595 GCGGACGTTTGGATTAAACGGG...CATTCGCTTTCGTTCCACCGTATT 1641
919 snAspValSerSerIleThrGlyLeuValAlaGlnSerAsnSerThrIle 935
1642 CAAATACCGATGAGAGGGCGATG..... 1665
936 IleAsnThrAspSerGlyIleIleAspLeuTyrGlyArgGlySerValG 952
1666ATTGTCAACCAATCAAGCAAGCAAGATCCACCGTTTACCA 1705
952 yMetLeuAlaIleAlaAspSerThrAlaGluAsnGlnGlyLysIleThrL 969
1706 TT.....ACGGCAATTAAGTAT 1725
969 euAspSerMetTrpValAspAlaAsnAspThrThrAlaMetArgAspIle 985
1726 GGTACA.....ACCGCAATTAACAGACGTGGA 1754
986 AlaSerAsnSerAlaIleAspPheGlyThrGlyValGlyValGlyThrAs 1002
1755 TACG.....AAAAAGAAATTCCTCAAC..... 1779
1002 pSerTyrSerGlyAlaGlyLysAsnAlaThrAlaIleAsnGlnLeuGlyG 1019
1780GGTGGCTTGGCGAG 1794
1019 lyValIleThrIleTyrAsnAlaGlyAlaGlyMetAlaIleTyrGlyAla 1035
1795 AAAGATACGACAAAGCAAGCGCGCTCAACCTTGT..... 1833
1036 SerAsnThrValIleAsnGlnGlyThrIleAsnLeuGlnLysAsnGlyAs 1052
1834 .TACCAGCCCGCGCAGACACCGCTCG.....CTCCTT 1870
1052 nTyrAspAspSerLeuAlaIleAsnThrLeuValGlyMetAlaValTyrG 1069
1871 CCGCGGCAACAATTTAAAGCGC.....AACATCAGCA 1905
1069 IuhIstGlyThrAlaIleAsnAspGlnThrGlyValIleAsnIleAsnVal 1085
1906 ACAACGCAAACTGTTTTCAGC...GGCAGACCAACA..... 1941
1086 GlyThrGlyGlnAlaPheTyrAsnAspGlyThrGlyThrIleValAsnTy 1102
1942CCGACGCTACATC 1957
1102 rGlyThrIleCysThrPheGlyValCysGlnSerGlyAsnGluTyrAsnA 1119


```
1958 ATTTAAAGCCATTGGTCGCAAAAAGAGGCAATTCCTCGCGGGAAATC 2007
      : : : : :
1119 snthraspsphetrserleu.....lle 1127
2008 GTGTGGAGACAGCTGATCAACCGCACATTAAAGCGGAAACTTCCA 2057
      : : : : :
1128 Tyrtthgllylaspthrargser.....glyluthrvalph 1142
2058 AATTAAAGCGGACAGCGGTGTTCCGCAATGTTGCCAAAGTGAAG 2107
      : : : : :
1142 rleuasnysserlalaivalthrapslyleualaglyasnvalaia 1159
2108 GCGATTGGCATTTGAGCAATCAAGCCCAAGCAATTTTGGTTCGACCG 2157
      : : : : :
1159 snserllythrleuSerGlyaspGlnlleThrValser..... 1171
2158 CATCAAGCCACACATCTGTACAGTTGAGCTGAGCGGTCTGACAA 2207
      : : : : :
1172 .....SerGlyleuLeuG1 1176
2208 TTGTGTGAAAAAACCATTAACCGACGATTAAGTGAATGCTTCATTGAC 2257
      : : : : :
1176 u.....AsnThrSerGlylylleleasnasnleuVal 1188
2258 AGACGAC.....ATCAGCGGCAAT 2277
      : : : : :
1188 ysluaspslysglyalavalilelysasnlaaglyvalmethrasn 1204
2278 GTGATCTTCGCGATCAGCGCTCATTTAAATCTCACAGGCGTTCGACACT 2327
      : : : : :
1205 valasprvalsergly...glylleuasnasnlaaglyglumethrAl 1220
2328 CAACGGCAATCTTACGCAATAGGCAATACAGTTATACAGTACGCCACA 2377
      : : : : :
1220 aglnlethrmetasnlaaglyalaspserSer...leuvalasnAnt 1236
2378 ACGCCCAAAAGCGCAACCTTACGCTTCGTGGGCAATGCCACACACA 2427
      : : : : :
1236 hnglyThrleasn.....LysllevalGlnasnlaaglyalphe 1249
2428 TTTAATCAAGCCATTAACCGCAACACATCGGCTTCGGCAATGCTTC 2477
      : : : : :
1250 AsnasnserGlySerValThrGlyArghmetsetrAlaglyGlyValph 1266
2478 ATTATCTAAGCCGACCGCCGATCAAAAGCGCAGTGTGACGCTTCG 2527
      : : : : :
1266 easnasnglnthrapsGlyalalemetArGlyala...Alaleuthrg 1282
2528 GCAACGCTAAGCA..... 2541
      : : : : :
1282 lyThrAlaValAlaasnasncluglyThrTPrasnleuglySerSerSer 1298
2542 .....AACGTAAAGCATTCGCGACTCAACGGTAATGTCTCCCTGAC 2582
      : : : : :
1299 gluglyAsnasnThrGlyMetleuglyValasnasnserAlaAlaAla 1315
2583 CGATTAAGCCAGTATTCATTTTGAAGACGCCCTTTACCGGACAAATCA 2632
      : : : : :
1315 nasnArGlyglupheilleuaspsnaspslys..... 1326
2633 GCGGCGGACAGATACGCGATTACCTTAAGACAGAGCAATGACGCTG 2682
      : : : : :
1327 .....AsnAlaValAlaHisleasnGln..... 1333
2683 CCGTCAGCGACG.....GAATTAGGCAATTTAAACCTTGACAAAGCCAC 2726
      : : : : :
1334 ...SerGlyThrleuTyAsnThrGlyHismetasnleSerasnSerSe 1349
2727 C..... 2727
      : : : : :
1349 rHisasncllyalavalasnmetTPrglyGlyAsncllyArghpheilleasn 1366
2728 .....ATTACACTCAATTC 2742
      : : : : :
1366 spGlyThrIleaspValSerAlaLysSerleuValValSerAlaasn 1382
      : : : : :
2743 GCC..... 2745
      : : : : :
1383 AlaGlyaspGlnasnAlaPhePheThrpsnclnaspsncllyleas 1399
2746 .TATCGCACATGCGCA..... 2763
      : : : : :
1399 nPheaspHispsrserAlaSerAlaValLysValThrHisSerAsnPhel 1416
2764 .....GGGCGCAACCGGCAAGTGGCGACAGTTCGCGCGC 2799
      : : : : :
1416 leaIaGlnasnaspGlyllemetasnleSerGlyThrGlyAlaValAla 1432
2800 GCGCGTTCGCGCGCGTTCGCGCGCTTCCTATATATCCGTTACCGGCCAAC 2849
      : : : : :
1433 MetGluGlyaspLysAsnlaaglnleuValasnasncllyThrIleasnle 1449
2850 TTCGTTAGAAATCCCGTTTCAACACAGCTGACGCGTAAACGCAATTGACG 2899
      : : : : :
1449 uGlyThrAlaGlyThrThrapsThrGlyMetlleGlymetGlnleuaspa 1466
2900 GTCAAGGACATTCGCGCTTATGTGGAACCTTCGCGC..... 2937
      : : : : :
1466 laasnAlaThrAlaaspAlaValllecluasnasncllyThrIleasnIle 1482
2938 TACCGCACCGCAAAATTGAAGCTGCGGAAAGTTCGACGACCTTACAC 2987
      : : : : :
1483 PheAlaasnAsp.....SerPheAlaPheSerValleuGlyThrValG1 1497
2988 CTTGCGCGTCAACAT..... 3003
      : : : : :
1497 yHisValValasnasncllyThrValVallleAlaaspGlyValThrGlys 1514
3004 .....ACGCGC 3009
      : : : : :
1514 erglyleuIlelysglnlyAspsertleasnvalGluGlymetasnGly 1530
3010 ACGCAACCTCAAGCCTCGAA..... 3030
      : : : : :
1531 AsnasncllyasnSerSerGluValHisThrGlyAspTyThrleuProAs 1547
3031 .....CAATTGACGATGAGGAAAGAAAGACAAACAAC 3064
      : : : : :
1547 pValProlyProAsnThrValSerValThrSerGlySerAspGlnAlaG 1564
3065 CGCTGTCC...GAAACTTAATTTCACCTCGCAAAAGCAACAGTGCAT 3111
      : : : : :
1564 lyGlySerMetasnAsnleuasncllyTyValValGlyThrAsnValasn 1580
3112 GCGCGCGCGTGGCGTTACCAACATCCGCAAGAGCGGAGTTCGCGCT 3161
      : : : : :
1581 GlySerAla.....GlyLysleuLysVa 1588
      : : : : :
3162 GCATTAAT...CCGTCAAAGACAAAGAGCTTTCGCAAACTCGGCAAG 3208
      : : : : :
1588 laasnAlaSerMetasncllyValGluIleasnThrGlyPheThrAlaG 1605
3209 CAGAAAGCAAAACAGCGCGGAAAAAGACAGCGGCAAAAGCTTGACGCG 3258
      : : : : :
1605 lyThrAlaAspThrThrValSerPheAspAsnValValGluGlySerAsn 1621
3259 CTGATTGCGCGCGCGCGATGCGCGTTCGAAAGAAAGAGGTT..... 3303
      : : : : :
1622 leuThrAspAla.....AspAlaIleThrSerThrSerValValThrPh 1636
3304 .GCCGAACGCGCGCGCGACGACGCGGGAATAATGTCGCGCATTAAGCAG 3352
      : : : : :
1636 rAlatysGlySerThrAspAlaSerGly...AsnValAspValThrMetS 1652
3353 CGGAGGAAGAAAAACGGGTGCGAGCGGATTAAGACACCGCTTGGCG 3402
      : : : : :
```



```
2638 GCGAAGATACGGCATTTACACTTA.....AAAGACAGCAATGGAC 2678
|||||
1300 GlyGlnAsnThrIleIleAsnLeuThrArgAlaAsnSperLeuThrG1 1316
2679 GCTGCGGTCAGGCGACGGAAATTAAGCAATTAACCTTGACACGCCACCA 2728
|||||
1316 nValAsnSerGlyThr...IleThrAlaThrAsnGlyTyrSerAlaIleG 1332
2729 TTACATCAATTCGGCC.....TATGCCACGATGGCGCA 2763
|||||
1332 hThrValAsnGlySerAsnSperProLysTrpIleThrAsnThrAlaThr 1348
2764 GGGCGCAAAACGGCGACGTGGACAGATGCGCGCGCGCTGGCGCGG 2813
|||||
1349 GlyValIleAsnGlyIleAsnProAspAlaPro..... 1359
2814 TTGCGCGCGTCCCTATATCCGTTACCGCCAACTTGGTAGAATGCC 2863
|||||
1360 .....LeuIleAsnLeu.....GlyA 1365
2864 GTTTCACACAGCTGACGCGTAAACGCAAAATTGAACGTCAGGAACTTC 2913
|||||
1365 rGgLyTyraSnpheGlyAsnGlnGlyThrIleAsnValGlnGly..... 1379
2914 CGCTTATGTGCGAACTTCGCGTACCGACGCAAAATTGAAGCTGGC 2963
|||||
1380 .....AspAsnAlaValAlaIle 1385
2964 GGAAGTTCGGAAGCACTTACACTTGGCGGTCACCAATACCGCAACG 3013
|||||
1385 eSerGlyGlyThrSerSerTyrValIleAsnLeuValAsnSerGlyThrI 1402
3014 AACCTGCAGCCTCGAACA..... 3033
1402 leAsnValGlyThrGlnGlnGlyLysGluAspGlyThrAsnGlyThrGly 1418
3034 TTACGCGTGTGGAAGAAAGCAAAACGCGTGTCCGAAAACCTTAA 3083
|||||
1419 LeuIleGlyIleLysGlyAsnGlyAsnAlaThrThrIleAsnAsnThrAl 1435
3084 TTTACACCTCGCAAAAGCAACAGCTGCAT.....GCCGGCGGTGGC 3124
|||||
1435 aAspGlyValIleAsnValTyrAlaAspAspSerTyrAlaPheGlyGlyL 1452
3125 GTTACCAACTCATCCGCAAAAGCGCGAGTTCGCGCTG..... 3162
|||||
1452 yThrLysAlaIleIleAsnAsnGlyGluIleAsnLeuLeuCyAspSer 1468
3163 .....CATAA 3167
1469 GlyCyAspIleTyrAlaProGlyThrGlyThrGlnAsnAspHisAs 1485
3168 T..... 3168
1485 nGlyThrAlaAspIleValIleProAspAlaThrThrAlaProThrGluG 1502
3169 .....CCGCTCAAAAGCAACAGCTTTC 3192
1502 ySerIleProThrProProAlaAspProAsnAlaProGlnGlnLeuSer 1518
3193 GACAACATC.....GGCAAGGAGAAAGC 3215
|||||
1519 AsnTyrIleValGlyThrAsnAlaAspGlySerSerGlyThrLeuLysAl 1535
3216 CAAAAAAGCGCGAAAAAGCAACGCGCAAGCTTGACGCGCTGATTG 3265
|||||
1535 aAsnAsnLeuValIleGlyAspAsnValLysValAspThrGlyPheThrS 1552
3266 GCGCGCGCGGATGCGCGTGAAGAAAGCAAGACGTTGCCGACCGGCC 3315
|||||
1552 eGlyThrAlaAspThrThrValValValAspAsnAlaPheThrGlySer 1568

3316 CGCAGCGACGCGGGGAAATGTC.....GGCATTAATGACGCGGA 3356
|||||
1569 AsnIleGlnGlyAlaAspAsnIleThrSerThrSerValValTrpAsnAl 1585
3357 GGAAGACAAAAACGGGTGACGCGGATTAAGACACCGCTTGCGCAAC 3406
|||||
1585 aGlnGlySerGlnAspAlaAspGlyAsnValAspValThrMetThrLysA 1602
3407 AGCGCAAGCGGAACCGCGCGGTACACCGCTTCCCGCGCGCGCC 3456
|||||
1602 snAlaTyrAlaAspValAlaIleThrAspSerSerVal..... 1613
3457 GCGCGCGCGCGGATTGCGCACTGCAACCCCAACCGCACCGCCAMCC 3506
|||||
1614 .....SerAspValAlaGlnAla..... 1619
3507 GCAAGCGCATGATCAACCGCTTATGCCAATAGCGTTG..... 3546
|||||
1620 .....LeuAspAlaGlyTyrThrAsnAsnGlnLeuTyrThrSerL 1633
3547 .....AGTAATTTTCCGCGACGCTCAACAGCGTTTC 3579
|||||
1633 euAsnValGlyThrThrAlaGluLeuAsnSerAlaLeuLysGlnValSer 1649
3580 GCGGTACAGGACGAATTAGACCGCGTATTGGCGAAGACCGC..... 3621
|||||
1650 GlyAlaGln.....AlaThrThrValPheArgGluAlaIleValLeuSe 1664
3621 ..... 3621
1664 rAsnArgPheThrMetLeuAlaAspAlaAlaProGlnIleLysAspGlyL 1681
3622 .....CGCAACGCGCTTGGACACGCGCATCCG..... 3651
1681 euAlaPheAsnValAlaAlaLysGlyAspProArgAlaGluLeuGlyAsn 1697
3652 GACACCAACACACTACCGTTGCGCAAGTTTCCGCGCTACCGCAACAAAC 3701
|||||
1698 AspThrGlnTyr.....AspMetLeuAlaLeuArgGlnThrLe 1710
3702 CGACCTGCC.....CAATTCGTATGCAAGAAA 3730
|||||
1710 uAspLeuThrAlaSerGlnAsnLeuThrLeuGlnTyrGlyLeuAlaArgL 1727
3731 ACCTGGCGGCGGCGCGTCCGATCCTGTTTGGCACACCGACCGAA 3780
|||||
1727 euAspLysAspLys.....SerLysThrAlaGlyAsp 1737
3781 AACACCTTCGACGACGCGATCGCAACTGCGACGCGCTTGCCGACGCGC 3830
|||||
1738 AsnGlyLeuThrGlyGlyTyrSerGlnPhePheGlyLeuLysHisSerMe 1754
3831 CGTTTTGGGGCAA.....T 3844
1754 tAlaPheAspGlnGlyLeuAlaIleTrpAsnAsnSerLeuArgTyrAspValH 1771
3845 ACGGCATGACAGGTTTACATCGGCATCAGCGCGCGCGGCTTTTACG 3894
|||||
1771 tAsnLeuAsp.....Ser 1775
3895 AGCGGACGCTTTCAGACGCGCATCGAGCAAAATC..... 3930
|||||
1776 SerArgSerValAlaTyrGlyAspValAsnLysIleLeuAspSerSpmE 1792
3931 .CGCGCGCGGTGCTCATTC.....GCATTCAGCACGATTAC 3970
|||||
1792 tArgGlnIleTyrLeuGlnIlePheArgSerGlnGlyAlaLysThrPheThrM 1809
3971 GCGCGGTTTCGCGGATTCGCGATCGCAACCGCAACGCGCGCAACGCC 4020
|||||
1809 eMetGly.....AspAlaLeuLysValIleThrProTyrAlaGlyLys... 1823
4021 TATTTCGTCAAAAAGCGGATTAACCGCTACGAAAGCGTAATATGCGCAC 4070
```

```

1824 ...PheargHisPheMetGluSpGlyTyrLysGluArgSerAlaGlyAs 1839
4071 CCCCCGCTGCATTCACCGCGTACCGC.....G 4099
1839 pPheasnLeuSerMetasnSerGlyasnGluAlaValaLaspSerIleV 1856
4100 CGGGCATTTAGCGAGATTATTCATTCAACCGCGCACACATTTCATC 4149
1856 alGlyLeuLysLeuAspTyrAlaGlyLysAspGlyTyrPserAlaThrAla 1872
4150 ACG.....CCTATTAGACCTGCTCCTATACCATGCGCGCTTC 4187
1873 ThrLeuGluGlyGlyProasnLeuSerTyrSerLysSerGln..... 1886
4188 GGGCAAGTCGCAACACGCGCTCAATACCGCGCTATTGCTCAGATTTCG 4237
1887 .....ArgThrAlaSerLeuGlnGlyAlaAlaGlyGlnSerPheG 1900
4238 GCAAAACCCGACGTCGGGAATGGGGCGTAAACCGCAATCAAGGTTTC 4287
1900 LyValaAspAspGlyGlnLysGlyGly..... 1908
4288 ACGGTGTCCTCCACGCTGCGCGCGCAACGCGCAACGTGAGACGCA 4337
1909 .....GlyValaAsnGlyLeuAlaThrIleGlyValLysTyrSerSerAs 1923
4338 ACACAGCGCGCGCATCAATTAATGCTACCGCTGG 4371
1923 nasPThrAlaLeuHisLeuAspAlaTyrGlnTyr 1934

seq_name: SwissProt_40:OMP_B_RICCN

seq_documentation_block:
ID OMP_B_RICCN STANDARD: PRT: 1655 AA.
AC Q9KKA3: Q9KKA3: Q9KKA3:
DC 16-OCT-2001 (Rel. 40, last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DI 01-MAR-2002 (Rel. 41, last sequence update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (surface protein antigen) (cell surface-exposed protein (surface protein
DE antigen) (120 kDa outer membrane protein ompb); 32 kDa beta peptide).
GN OMPB OR RCI085.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Sanson D., Roux V., Cossart P., Weissbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
RN [2]
RP SEQUENCE OF 33-1649 FROM N.A.
RC STRAIN=Indian tick typhus; and Malish 7;
RX MEDLINE=20393643; PubMed=10939649;
RA Roux V., Raoult D.;
RT "Phylogenetic analysis of members of the genus Rickettsia using the
RT gene coding the outer membrane protein ompb (ompb).";
RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
RN [3]
RP SEQUENCE OF 353-1655 FROM N.A.
RC STRAIN=Malish 7;
RA Stenos J., Walker D.;
RT "The rickettsial outer membrane protein A and B genes of Rickettsia
RT australis, the most divergent rickettsia of the spotted fever group.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR

```

```

CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC SIMILARITY).
CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AE008659; AAL03623.1; -
DR EMBL, AF123721; AAF34124.1; -
DR EMBL, AF123726; AAF34129.1; -
DR EMBL, AF149110; AAD39533.1; -
DR Interpro: IPR03858; rompa_ompb.
DR Pfam: PF02708; rompa_ompb.1.
KW Antigen; S-layer; Cell wall; Complete proteome.
FT CHAIN 1 1334 120 KDA SURFACE-EXPOSED PROTEIN.
FT VARIAT 1335 1655 32 KDA BETA PEPTIDE.
FT VARIAT 61 61 P -> A (IN STRAIN INDIAN TICK TYPHUS).
FT VARIAT 75 75 G -> S (IN STRAIN INDIAN TICK TYPHUS).
FT VARIAT 78 78 K -> N (IN STRAIN INDIAN TICK TYPHUS).
FT VARIAT 251 251 V -> A (IN STRAIN INDIAN TICK TYPHUS).
FT VARIAT 413 413 N -> D (IN STRAIN INDIAN TICK TYPHUS).
FT VARIAT 959 959 I -> V (IN STRAIN INDIAN TICK TYPHUS).
FT VARIAT 988 988 A -> T (IN STRAIN INDIAN TICK TYPHUS).
FT VARIAT 1139 1139 R -> L (IN STRAIN INDIAN TICK TYPHUS).
FT VARIAT 353 354 F -> S (IN REF. 3).
FT CONFLICT 776 776 E -> S (IN REF. 3).
FT CONFLICT 1159 1159 G -> S (IN REF. 3).
FT CONFLICT 1177 1177 G -> S (IN REF. 3).
FT CONFLICT 1492 1492 H -> R (IN REF. 3).
FT SPOUNCE 1655 AA; 168342 MW; E49E1937D5FC37 CRC64;

alignment_scores:
Quality: 245.00 Length: 1437
Ratio: 0.364 Gaps: 64
Percent Similarity: 46.834 Percent Identity: 20.251

alignment_block:
US-09-303-518D-649 x OMP_B_RICCN ..
Align seg 1/1 to: OMP_B_RICCN from: 1 to: 1655

679 CTCGTGTGTCACATACCTTTGCAACAAATGAGACAGTGGGACAGT 728
:::||||| ||||| ||||| ||||| ||||| |||||
199 ITeLguglyAspAsnAlaValIleValAsnGlyAla...AsnGlyThrIle 214
729 CAACCTAGGTAGTGAATAATTAAACATGACCCATGCTTTTATACAA 778
:::||||| ||||| ||||| ||||| ||||| |||||
214 uasnValThrAsnGlyPheIleLysValSerSerLysPheAsnThrIle 231
779 CAGGAGGC...TCATTGGCGACAGTGGCTCACCATAATGTTATCATGAT 825
:::||||| ||||| ||||| ||||| ||||| |||||
231 alAsnValIleAsnIleGlyAspGlyGlnGlyIleMetPheAsnThrIle 247
826 GCCCAAAAGCAAAAGTGTGA..... 846
|||:::||||| ||||| ||||| ||||| ||||| |||||
248 AlaAspAsnValAsnThrLeuAsnLeuGlnAlaAsnGlyAlaThrIle 264
847 ATTATAGGATATTGCAACAGCGCAACCCCTATATAGCAAAAGCATG 895
:::||||| ||||| ||||| ||||| ||||| |||||
264 rPheasnGlyThrAspGlyThrGlyArgLeuValLeuSerLysAsnA 281
896 GCTTCCAGCTGCTGCTAAAGATTTGTTCTATGATGAATCTTTGCTGCA 945

```

```
281 Ia.....AlaIaIaThrasPheasnValThrGlySerLeuGlyGly 294
946 GATACCATTCAGTA.....TTCTAGCAACACAGCTCAAAATGGGAATA 999
295 AsnLeuLysGlyIleIleGluPheasnThrValAlaValasnGlyGlnLe 311
990 CTCTTTTAAC.....GACGATATAATAGGCA 1015
311 ULysAlasnAlaGlyAlaAlaAlaValIleGlyThrAsnGlnGlyA 328
1016 CAGGAAATATCAATGCCAACAATGACACAAATTCCTGCTTAAT...AGA 1062
328 IaGlyArg.....AlaIaGlyPheValValSerValaAspAsnGlyLys 342
1063 TTAACACGACGACGCTTCAATGTTTAAATGTTCTTTATCCGACAGC 1112
343 ValAlaThrIleAspGlyGlnValTyr.....Al 352
1113 AAGAGAACCTGTTATATCATGCTGCAGGTGTCAC.....AGTTATC 1156
352 aLysAspMetValIleGlnSerAlaAsnAlaValIleGlnValaAsnPheA 369
1157 GACCCAGACTGAATATGSA...GAAATATTCCTTTATTCACGACGA 1203
369 rGHisIleValaAspValGlyThrAspGlyThrAlaPheLysThrAla 385
1204 AAGGCGAATGTGATCTACTACGACACATCATCAAGTGT..... 1245
386 AlaserLysValAlaIleThrGlnAsnSerAsnPheGlyThrThrAspH 402
1246 GGAGGATTTATTCCAA.....G 1264
402 eGlyAsnLeuAlaIleGlnIleLeValProasnThrMetThrLeuAsnG 419
1265 GAGATTTTACGTCGTGCGCTGAATAAGCAAACTTGCAAGCGCGCGG 1314
419 LysAsnPheThrGlyAspAlaSerAsnProGlyAsnThrAlaGly..... 433
1315 GTTCATATCATGTCAGACAGTACCGTTACTTGAAATGAAGCGC..... 1359
434 .....ValIleThrPheAspAlaAsnGlyThrIle 443
1360 .....GTGCAACGACCGCCCTGT 1378
443 uAlaSerAlaSerAlaAspAlaAsnValAlaValThrAsnAsnIleThrA 460
1379 CCAAAATCGCAAAAGCGCTGCACGTTCAAGCCAAAGGCGAA..... 1422
460 IaIleGluLaserGlyAlaGlyValaGlnLeuSerGlyThrHisAla 476
1423 .....AACGAAGCTGATC...AGCTGGCGGACGG 1451
477 AlaGluLeuArgLeuGlyAsnAlaGlySerValPheLysLeuAlaAspG 493
1452 TACAGTATTTTGATTCAGCAGCAGAGATTAAGCAAAACAAAGCGT 1501
493 yThrValIle.....AsnGlyLysValaAsnGlnT 503
1502 TTAGTGAATCGGCTTGCACGCGCAGGAGGTACGGTCAACTGAATGCC 1551
503 hAlaLeuValaGlyIleValaLeuAlaIaGlyThrIleThrLeuAspG 519
1552 GATATCAAGTTCAACCCGACAAACTGATTTGGGCTTTCGGCGGACG 1601
520 SerAlaThrIleThrGlyAsp.....IleGlyAsnAlaGlyGlyAl 533
1602 TTTGATTTAAAGCGCATTCGCTTGCCTTCCACCGTATTCAAAT...A 1648
533 aAlaAlaLeuGlnGlyIleThrLeuAlaAsnAspAlaThrLysThrLeuT 550
1649 CCGATGAAGGCGCATGTGTCAAACCAAT..... 1680
```

```
550 hLeuGlyGlyAlaAsnIleIleGlyAlaAsnGlyGlyThrIleAsnPhe 566
1681 CAAGCAAGAAGATCCACGTTACCATTAAGCAAGCAATTAAGAT..... 1722
567 GlnAlaAsnGlyGlyThrIleLysLeuThrSerThrGlnAsnIleVala 583
1723 .....ATTCGTACA.....ACCGCG.....ATA 1741
583 lValaAspPheAspLeuAlaIleAlaThrAspGlnThrGlyValaValaAspA 600
1742 ACAACAGCTGGATGATGCAAAAAAGAAATTCCTTACACAGCGTTGTTGCC 1791
600 lAserSerLeuThrAsnAlaGlnThrLeuThrIleAsnGlyLysIleGly 616
1792 GAGAAAGAT...ACGACCAAAACGACGCGGCTCAACCTGTTTACCA 1838
617 ThrValaGlyAlaAsnAsnLysThrLeuGlyGlnPheAsnIle..... 630
1839 GCCCGCGCGAAGACCGCACCTGCTGCTTCGCGGACGACAAATTTAA 1888
631 .....GlySerSerLysThrValaLeuSerAspGlyAspValaAlaIleA 645
1889 ACGCAACATCAACGCAAAACGCAAACTGTTTTCAGCGGCGACAGCA 1938
645 snGluLeuValIleGlyAsnAsnGlyAlaValaGlnPhe..... 657
1939 ACACCGCACCGCTACAT...CATTTAAACGACCATTTGTCGCAAAAGA 1985
658 .....AlaHisAsnThrTyLeuIleThrArgThrThrAsnAlaAl 671
1986 GGCATTCCTCGCGGGGAAATCGTGGGACACACAGCTGATCAACCGCA 2035
671 aGly.....GlnGlyLysIleIlePheAsnProValaValaAsnAsnA 686
2036 CATTTAAAGCGGAA..... 2049
686 hThrIleuAlaThrGlyThrAsnLeuGlySerAlaThrAsnProLeuAla 702
2050 .....ACTTCCAATTTAAAGCGGACAGCGCGTG...GTTTCCGCGCA 2090
703 GluIleAsnPheGlySerLysGlyAlaAlaAsnValaAspThrValaLeuAs 719
2091 TGTTCGCAAGTGAAAGCGCATTTGGCATTTGACAAATCACGCCACAGAG 2140
719 nValaGlyLys..... 722
2141 TTTTGGTGTGCGACCGCATCAAGCCACACAAATCTGTACAGTTGCGAC 2190
723 .....GlyValaAsnLeuTyralaThrAsnIleThrThrAspAlaAsn 737
2191 TGG.....ACGGGCTGCAAAATTTGTGCGAAAAAC 2222
738 ValGlySerPheIlePheAsnAlaGlyIleThrAsnIleValaSerGlyThr 754
2223 CAT.....ACCGAG 2233
754 rValaGlyGlnGlnGlnGlyAsnLysPheAsnThrValaAlaLeuAspAsnG 771
2234 ATAAAGTATGCTCATTTGACTAAGACCGACATCAAGCGCATGTGCAT 2283
771 TyThrThrValaLysPheLeuLysAlaThrPheAsnGlyAsnThrThr 787
2284 CTTCGCGATCAACGCTCAATTAATCTCACAGGCTGCTCCACACATCAAGC 2333
788 lIleAlaAlaAsnSerThrLeuGlnIleGlyLysnTyThrAlaAspH 804
2334 CAATCTTATGCAATGCGCATACACTTATACGTACGTACGACACACACCA 2383
804 eValaAlaSerAlaAspGly...ThrGlyIleValaGluPheValaAsnThrGly 820
2384 CCAAAAGCGCAACCTTAGCTCTGGGCAATGGCCAAAGCAACATTAAT 2432
821 ProIleThrValaThrLeuAsnLysGlnAlaAlaProValaAsnAlaLeuTy 837
```



```

1347 LUMETALAGLY..ProGluAlaGlyAlaLeuProAlaAlaValAlaAlaG 1363
4037 CGGATTACCGGTACGAAACATCAATATGCCACCGCCGCTTGCAATTC 4086
1363 LysAspGluAlaValAlaAspAsnValAlaTyrGlyIleTyrAlaLysProPhe 1379
4087 .....AACCGCTACCGCGCGCATTAAGCAGATATTCATT 4124
1380 TyrThrAspAlaHisGlnSerLysGlyGlyLeuAlaIleTyrLysAl 1396
4125 CAACCGCGCG.....AACCGCTACCGCGCATTAAGCAGATATTCATT 4141
1396 aLysThrThrGlyValAlaIleGlyLeuAspThrLeuAlaAsnAspAsnL 1413
4142 TTTCATCAGCGCTTATTTGAGCGCTGATACCGATACCGCGCTTGCGGC 4191
1413 eumETileGlyAlaAlaIleGlyIleThrLysThrAsp..... 1425
4192 AAAGTCGGAACACCGCTCAATACCGCGCTATGCTCAGGATTTGCGCA 4241
1426 .....IleLysHisGlnAspTyrLysLys 1433
4242 AACCCGAGTGGCGAATGGCGCTAAACGCCGAATCAAGTTTCACGC 4291
1433 S.....GlyAspLysThrAspValAlaAsnGlyPheSerP 1444
4292 TGTCCCTCAGCGCTGCGCGCGCAAGCGCGCACTGAGACGCGCAAC 4341
1444 heserLeuTyrGlyAlaGlnGlnLeuValLysAsnPhelaGlnGly 1460
4342 AGCGCG 4347
1461 SerAla 1462

seq_name: SwissProt_40:OMPA_RICRI
seq_documentation_block:
ID OMPA_RICRI STANDARD: PRT: 2249 AA.
AC P15921:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein A precursor (190 kDa antigen) (cell surface
GN OMPA.
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_taxid=783;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-R:
RX MEDLINE=90354033; PubMed=2117566;
RA Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;
RT "A protective protein antigen of Rickettsia rickettsii has tandemly
RT repeated, near-identical sequences.";
RL Infect. Immun. 58:2760-2769(1990)
CC -1- FUNCTION: ELICITS PROTECTIVE IMMUNITY.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC -1- SLAYER WITH HEXAGONAL SYMMETRY.
CC -1- PPM: GLYCOSYLATED (PROBABLY).
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M31227; AAA26380.1; -

```

```

DR PIR; A41477; A41477.
DR InterPro; IPR003858; rompa_rompb.
DR Pfam; PF02708; rompa_rompb; 1.
KW Antigen; Repeat; Signal; Cell wall; S-layer; glycoprotein.
FT SIGNAL 1 28
FT CHAIN 1 28
FT DOMAIN 29 2249
FT REPEAT 212 1180
FT REPEAT 212 1180
FT REPEAT 287 358
FT REPEAT 359 430
FT REPEAT 431 505
FT REPEAT 506 577
FT REPEAT 578 652
FT REPEAT 653 724
FT REPEAT 725 799
FT REPEAT 800 874
FT REPEAT 875 949
FT REPEAT 950 1021
FT REPEAT 1022 1093
FT REPEAT 1094 1165
FT REPEAT 1166 1180
FT SEQUENCE 2249 AA; 224333 MW; A9D664C089DF087 CRC64;

```

```

alignment_scores:
Quality: 242.50 Length: 1499
Ratio: 0.357 Gaps: 75
Percent Similarity: 45.364 Percent Identity: 20.080

```

alignment_block:

US-09-303-518D-649 x OMPA_RICRI

Align seg 1/1 to: OMPA_RICRI from: 1 to: 2249

```

655 CATATTGCAAGTGGTATTCCTGCTGTGCT.....GGCAATAC... 696
938 AsnLeuAsnGlyAlaLeuSerGlnValThrGlyAspIleGlyAsnThrAs 954
697 .....TTTGACAAATATGATCAGTGGTGGCACACTCACTAGTAGTAGT 742
954 nSerLeuAlaThrIleSerValGlyAlaGlyThrAlaThrLeuGlyGly 971
743 AAAAATTAACATAGCCCATATGCTTTTATCCAAAGGAGGCTCATTT 792
971 lValAlaLys..... 974
793 GCGGACAGTGGCTCACCAATGTTATCTATGATGCCAAAGCAAAAGTG 842
975 .....AlaThrThrLysLeuThrAspAlaLysAlaValLys 988
843 GTTAATTAATGGGTATGCAACCGCGCAC.....CCGATATACGAAAA 889
988 spherThrAsnProValAlaValThrGlyAlaIleAspAsnThrLysAsn 1005
890 GCAATGCTTCACGCTGCTGTAAGATGTTCTATGATGAATCTTT 939
1005 lAsn..... 1006
940 GCTGAGATACCATTCATGATTTCTAGCAACACGCTCAAAATGGGAATA 989
1007 ..... 1010
990 CTCCTTTAAGCAGATTAATTAATGACAGAGAAATCAATGCCAAACATG 1039
1010 lThrThrGlyAsnSerThrValThrGlyAsnValGly.....A 1024
1040 AACCAATTCCTGCT..... 1056
1024 snThrAsnAlaLeuAlaThrValAsnValGlyAlaGlyLeuLeuGlnVal 1040
1057 .....ATAGATTAAACAGCAACGCTCAATGTTATGTTCTTT 1100
1041 GlnGlyGlyValValLysAlaAsnThrIleAsnLeuThrAspAsnLase 1057

```


1635 yserValserglyValglyThrAlaAlaAsnSerValGlyAspIle 1652
2402 GCCGTGGGCG.....AATGCCCAAGCACA 2427
1652 hrThAlaGlyAlaThrSerPheAlaSerSerValAsnAlaLysGlyThr 1668
2428 TTTAATCAAGCCACATTAAACGCGACACATCGCTTCGGGCAATGCTTC 2477
1669AlaThrLeuGlyGlyThrThrSer..... 1676
2478 ATTAAATAGCGACCGCCGACAAACGCGAGCTGACGCTTCCG 2527
1677PheAlaAsnThrPheThrAsnThrGlyAlaValThrLeu.... 1689
2528 GCAACGCTAAGGCAACGATTCGCGCATTCGCAACGCTAATGCTTCG 2577
1690AlaLysGlySerIleThrSerPheAlaLysAsnValThrAlaThr 1704
2578CTAGCCGATTAAGCGCATTCATTTTGAAGCAGC...CGCTT 2618
1705 SerPheValAlaAsnSerAlaThrIleAsnPheSerAsnSerLeuAlaPh 1721
2619 TACCGGCAAAATCAGCGCGCGCAAGATAGCGATTACACTTAAAGACA 2668
1721 eAsnSerAsnIleThrGlyGly..... 1728
2669 GCGAATGAGCGCTGCCGTGAGCGACGCAATTAAGCAATTTAAACCTTGAC 2718
1729GlyThrThrLeu...ThrLeuGlyAlaAsn 1737
2719 AACGCCACATTAACACTCAATTCGCTTATGCCACGATGCGGACGGGCG 2768
1738 GluValThrThr..... 1742
2769 GCAACCGCGAGTGCAGACATGCGCGCGCGCGCTTCGCGCGCTGCG 2818
1743 GlyThrGlySerPheThrAspThr..... 1750
2819 GCGGTTCCTATATCCGTTACACCGCCCACTTCGCTAGATCCGCTTTC 2868
1751LeuThrLeuAsnThrThrPheAspGlyAlaAlaLysSerGlyGly 1765
2869 AACACGCTACGCTGAACGCGCAAA.....TTGAACGCTGACGGGAC 2909
1766 AsnIleLeuIleLysSerGlySerThrLeuAsnLysSerGlyValSerTh 1782
2910 ATTCCGCTTATGTCGGAACCTTCGCTACCGCAGCGACAATTTGAAGC 2959
1782 rLeuAlaLeuValValThrAlaThrAsnPheAspMetAsnIle.... 1797
2960 TGGCGGAAGTTCGGAAGCACTTACACTTGGCGGTCAACAATACC... 3006
1798SerProAspThrIleThrValIleSerAlaGluThrAla 1811
3007GGCAACGAACTGCAAGCTTCGAACAATGACGTAAGGAAGAAA 3053
1812 GlyGlyLeuLysProThrSerLysGluAsnValLysIleThrIleAsnAs 1828
3054 AGACAACAACCGCTGTC.....GAAACCTTAATTTCCAGCTGC 3094
1828 nasPAsnArgPheValAsnPheThrPheAspAlaSerThrLeuThrLeuP 1845
3095 AAAACGAACACGTCGATCGCGCGCGGTTCACCAACTCATCCGAAA 3144
1845 heAlaGluAspIleAlaAlaAspVal.....Ile 1854
3145 GAGCGGAGTTCGCGCTGCATTAATCCGCTC..... 3174
1855 AspGlyAspPheAlaProGlyGlyProLeuAlaAsnIleProAsnAlaAl 1871
3175AAAGACAAGACTTTCGACAAACTCGCAAGGCAAG 3214

1871 AsnIleLysLysSerLeuGluLeuMetGluAspAlaProAsnGlySerA 1888
3215 CCAAAAAACGCGGAAAAAGAC.....AACGCGCAAGC 3249
1888 sPAlaArgIleAlaPheAsnAsnPheGlyLeuMetThrProLeuGlnGlu 1904
3250 CTGACGCGCTGATTCGCGCGCGCGGATCGCGTCAAAAGACACAAG 3299
1905 AlaAspAlaThrThrThrIleLeuIleGlnAspValValLysProSerAspTh 1921
3300 CGTTGCCGAACCGCGCGCGGACGAGCGGGAATGTCGCGATTATGC 3349
1921 rIleAlaAlaValAsnAsnGlnValValAlaSerAsnIle..... 1934
3350 AGCGGAGAGAAAGAAAAAGGTCGACGCGGATTAAGACACCGCTTCG 3399
1935SerSerAsnIleThrAlaLeu 1941
3400 GCGAACAACGCGGACGGAACCGCGCGCTACACCGCGCTTCCCGG 3449
1941 1941
3450 CGCGCGCGCGCGCGCGGATTTGCCGCACTG...CAACCCCAACGCG 3496
1942AsnAlaArgMetAspLysValGlnSerGlyAsnLysGlyProV 1956
3497 AGCCCCAACCGCGAGCGGACCTGATACCGCTTATGCCAATGACGCTTG 3546
1956 alSerSerGlyAspGluAspMetAspAlaLysPhe...GlyAlaThrIle 1971
3547 AGTGAATTT...TCCGCCACG.....CTCAACAGCGCTTTGCG 3581
1972 SerProPheValGlyAsnAlaThrGlnLysMetCysAsnSerIleSerG 1988
3582 GTTACAGACGAATTAAGCGCGGTATTTGCCGAAGACCGCGCAACGCGG 3631
1988 yThrLysSerAspThr..... 1993
3632 TTTGACAGCGGACCTCCGCGACACCAACACTACCTTCGCAAGATTTC 3681
1994ThrGlyThrIleGlyPheAspGlyPheValSerAspAspLeu 2008
3682 CGC.....GCTTAC...CGCCACAACCGCACTGCGCCCAATTCG 3719
2009 AlaLeuGlyLeuAlaThrThrArgAlaAspThrIleLys..... 2022
3720 TATGCAAAAAACCTCGCGACGCGCGCTCGGCAATCTGTTTCGCACA 3769
2023LeuLysAsnA 2026
3770 ACCGACCGAAAAACACTTCGACGACGCGCATCGCAACTCGCGACGCTT 3819
2026 snLysThr.....GlyAspLysAsnLysValGluSer 2036
3820 GCCCAGCGCGCGCTTTCGGCAATPACGCGATCGACAGGTTTCACATCG 3869
2037 AsnIleThrSerLeuGlyGlyLeuThrAsnValProThrGluAsnLeuP 2053
3870 CATCAGCGCGCGCGGCTTTCAGCAGCGGAGCCTTCAGACGCGATCG 3919
2053 eValGluAlaIleAlaSerLysSerAsp..... 2062
3920 GAGCAAAATCCGC.....CGCGCGTGCATTCAGCGCATTCAG 3960
2063AlaIleArgSerLysSerArgArgValIleAlaThrThrLeuGlu 2078
3961GCACGATACCGCGCGGCTTCGCGGATTCGCGCATTCGACCGCAAT 4007
2079 ThrValGlyThrGlnThrAlaAsnGlyLysThrLysSerGlySerThr 2095
4008 CGCGGCAACGCGCTATTTCTCCAAAAAGCGGATTCACCGCTAC..... 4050
2095 rGlyGlnLeu.....MetAlaGlyThrThrThrLysMetLeuP 2107

4051 ..GAAGAGTCAATATGCGACCCCGGCTTCATCAGCGTACGCG 4098
 |||||..... ||||| ||||| |||||
 2107 roGlnsnlleasnleu...trpPro...LeuAlaGlyLeuArgTyr... 2120
 4099 GCGGCGCATTAAGCAGATTAATTCATCAACCGGCG.....CA 4136
 ||||| ||||| ||||| ||||| |||||
 2121 SerThrIleLys...AspLysGlyTyrLysGluThrGlyThrTyrG1 2136
 4137 ACACATTTCCATCAGCCCTTATTGACCGCTGTCATACCGATGCCGTT 4186
 |||||..... |||||
 2136 naenleuthrValLysGlyLysasnTyrAsnThrPheaspGlyLeuLeuG 2153
 4187 CCGGCAAGTCCGACACCGCATATAC..... 4215
 |||||..... |||||
 2153 LysAlaLysValSerSerAsnIleAsnValAsnGluIleValLeuThrPro 2169
 4216GCCGTAATGCTCAGATTTCGCGCAACCCGACGTCGGA 4256
 |||||..... |||||
 2170 GluLeuTyrAlaMetValAspTyrAlaPheLysAsnLysValSerAla... 2185
 4257 ATGGGCGCTAAAGCCGCAATCAAGGTTTCAGCGTCTGCTCCACGCTG 4306
 |||||..... |||||
 2186IleaspLarGluLeuInolymethrAlaProLeuProThrA 2200
 4307 CCGCGCCCAAGCGCCGCAACTGGAAGCGCAACAGCGCGGCGATC 4353
 |||||..... |||||
 2200 snserPheLys.....GlnSerLysThrSerPheaspAlaGlyVal 2213

seq.name: SwissProt_40:YDEK_ECOLI

seq_documentation_block:

ID YDEK_ECOLI STANDARD: PRT: 1325 AA.

AC P32051: P76140; P77168;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical lipoprotein ydek precursor (ORF7).
 GN YDEK OR ORF7 OR B1510.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. II, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97251357; PubMed=9097039;
 RA Alta H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saio N.,
 RA Sampedo G., Seki Y., Sivasubramanian S., Yamaguchi H., Yakeda J.,
 RA Takeuchi K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.,
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).
 RN [3]
 RP SEQUENCE OF 595-1325 FROM N.A.
 RX MEDLINE=94100243; PubMed=8274505;
 RA Cartwright P.J., Timms M.W., Lithgow T., Hoef P.B., Hoogenraad N.J.;
 RT "An Escherichia coli gene showing a potential ancestral relationship
 to the genes for the mitochondrial import site proteins ISP42 and
 MOM38.";

RL Blochim. Biophys. Acta 1153:345-347(1993).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
 CC (potential).
 CC -1- SIMILARITY: TO E. COLI YDEK.
 CC -1- SIMILARITY: SOME, TO FUNGAL MITOCHONDRIAL IMPORT SITE PROTEINS
 CC ISP42 AND MOM38.
 CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
 CC FRAMESHIFT IN POSITION 653.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE000248; AAC74583.1; -.
 CC EMBL: D90793; BAA15190.1; ALT_INIT.
 CC EMBL: D90794; BAA15197.1; ALT_INIT.
 CC EMBL: X73295; CAA51730.1; ALT_FRAME.
 CC PIR: S34315; S34315.
 CC Ecogene: EG11780; ydek.
 CC DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 CC DR Hypothetical protein; Membrane; Lipoprotein; Signal;
 CC KW Complete proteome.
 CC FT SIGNAL 1 18 POTENTIAL.
 CC FT CHAIN 19 1325 HYPOTHEMETICAL LIPOPROTEIN YDEK.
 CC FT LIPID 19 19 N-ACYL DIGLYCERIDE (POTENTIAL).
 CC FT CONFLICT 884 884 M -> K (IN REF. 3).
 CC FT CONFLICT 1317 1317 M -> S (IN REF. 3).
 CC FT SEQUENCE 1325 AA; 136514 MW; 26A3A06FA19AD7D CRC64;

alignment_scores:

Quality: 240.50 Length: 1492

Ratio: 0.362 Gaps: 80

Percent Similarity: 44.504 Percent Identity: 20.040

alignment_block:

us-09-303-518d-649 x YDEK_ECOLI ..

Align seg 1/1 to: YDEK_ECOLI from: 1 to: 1325

619 GATGAGATGAGCCCAATTAACCGGAAAGTTCATAT...CATATGCAAG 665
 |||||..... |||||
 67 AspsnaspGlnIleThrAsnIleAspThrAspValAlaTyrSpAlaTy 83
 666 TGGCATTTCTTG...CTCGTTGGTGCA 691
 |||||..... |||||
 83 rleuValGlyTyrPyrGlyThrGlyValIleAsnIleLeuAlaGlyGly 100
 692 AT.....ACCTTGGCAAAATGATCA 714
 |||||..... |||||
 100 snAlaSerLeuThrThrIleThrThrSerValIleGlyAlaAsnGluAsp 116
 715 GGTGGTGGCAGACGCAACTTA.....GGTATGAAATAATTAACATAG 758
 |||||..... |||||
 117 SerGluGlyThrValAsnValLeuGlyGlyThrThrPyrGlyTyrAsp 133
 759 CCATATGATGTTTTTACCAACAGAGGCTCATTTGGCGACAGTGGCTCAC 808
 |||||..... |||||
 133 rGlysnAsnAlaArgProLeu.....AsnValGlyGlnSerGlyThrG 148
 809 CATGTTTATCTATATGATGCCCAAAAGAGTGGTAAATTAATGAGGTA 858
 |||||..... |||||
 148 lYThrLeuAsnIle.....LysGlnLysGlyHisValAspGly... 160
 859 TGGCAACGCGCAACCCCTATATAGAAAGCAATGCGTTCGAGCTGT 908
 |||||..... |||||
 161GlyTyrLeuArgLeuLysSerThrGly..... 170
 909 TCGTAAAGATTGTTCTATGATGAATCATCTTTCGCGAGATACCATTCAG 958

```

170 ..... 170
959 TATTCAGAACCGTCAAAATGGAATCTTTTAAACGAGATAT 1008
170 ..... 170
1009 AATGCGACAGGAAAAATCATGCCAATGACACATCTCTGCTAA 1058
171 ..GlyValGlyThrValAsnValGlyGlnAspSerValLeu... 184
1059 TAGATTAAACACAGACCGTTCAATTGTTTATGTT... 1095
185 .....ThrThrGluLeuPheGluIleGlySerTyrGly 196
1096 .....TCTTATCCGAGACAGACAGAACGCTTTTTCATGCTGCAGCT 1140
196 hrGlySerLeuAsnIleThrAspGlyTyrValThrSerSerIleVal 212
1141 GGTGTCAACAGTTATGCAACCCAGCTGAATATGAGAAAAATATTTCTT 1190
213 AlaIleLeuGlyTyrGlnAla.....GlySerAsnGlyGlnVal 225
1191 TATTGACGAGAAAGGCGAATTGATTAATTACCAACATC..... 1233
225 lValValGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 242
1234 .....AATCAAGTGTGCGAGATTAATATTTCCAAAGA 1266
242 leGluPheGlnIleGlyAsnGlnGlyThrGly..... 252
1267 GATTTTACGGCTCGCCTGAAATTAACGAACCTTGCAAGCGCGCGCT 1316
253 .....GluAlaThrIleArgGlyGlyGlyGlyGlyGlyGlyGlyGly 261
1317 TCATATCAGTGAAGACAGTACCTTTGTAAGAAATGAAGCGCGGCGCA 1366
261 uValThrAlaGlyAsnThrIleIleGlyGlyGlyGlyGlyGlyGlyGlyGly 276
1367 ACAGCGCGCTTCCAAAATCGCAAGGACGCTGACGCTTCAACCAAA 1416
277 .....GlyThrLeuAsnValGln..... 282
1417 GGGGAAACCAAGGCTCGATCAGCGT..... 1443
283 ...AspGlnAspSerValIleThrValArgArgLeuTyrAsnGlyTyrPhe 298
1444 GCGCAGGTACAGTCAATTTGGATCAGCAGGACAGATGAAGCAAAA 1492
298 eGlyAsnGlyThrValAsnIleSerAsnAsnGlyLeuIleAsnAsnGlyG 315
1493 AACAAAGCCTTGTGAATCGGCTTGCTGACGCG...AGGGTACGCTG 1539
315 lu.....TyrSerLeuValGlyValGlnAspGlySerHisGlyValVal 329
1540 CAAGTGAATCCGATATCAGTTCAACCCGACAAACCTATTTCCGGCTT 1595
330 AsnValThr.....AspGlySerGlyHisThrPasnPh 339
1590 TCGCGCGCGAGCTTGATTTAAAGGGCATGCTTTCCTCCACCGCTA 1639
339 eLeuGly.....ThrGlyGlnAlaPheArgTyrIleTyrI 351
1640 TTCAAAATACGATGAAGGGCGGATGTGCAACCCACATCAAGACAAA 1689
351 leGlyAspAlaGlyAspGlyGlyLeuAsnValSerSerGlyGlyValVal 367
1690 GAATTCACCGTTACCATTTACAGCAATATATGCTCAACGCGCAAA 1739
368 AspSerGlyIleIleThrAlaGlyMetGlyGlyGlyGlyGlyGlyGlyGlyGly 383
1740 TTAACAACGCTTGATAGCAAAAAAATGCTTACACGCTTGCTTG 1789

```

```

383 nIleThrValLysAsp..... 388
1790 GCGAAGAAATATGACCAACCAAGCGCGCTCAACCTGTTTACAG 1839
389 .....LysAsnSerValIleThrAsnLeuGlyThrAsnLeuGlyTyrAsp 403
1840 CCCGCCGAGAA.....GACCGACCGCTGCTGCTTCC...GG 1874
404 GlyHisGlyGlyLeuMetAsnIleSerAsnGlnGlyLeuValValSerAsnG 420
1875 CGGACCAATTTAAAC.....GGCACATCAGCGCAAA 1906
420 yGlySerSerLeuGlyTyrGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 437
1907 CAACGCGCAACGTTTTCAGGCGCGACACACACCGCGCTTCAAT 1956
437 hrThrGlyGlyMet..... 441
1957 CATTTAAACGACATTGTCGCAAAAAGCGCATTCGCGGGGAAAT 2006
441 ..... 441
2007 CGTGGGACACGACGATGATCAACCGCACATTTAAAGCGAA..... 2049
442 ...TyrPheGlu.....ValAsnLysAsnValTyrThrThrIleGlyVal 454
2050 .....AATTCGCAATTTAAAGCGGAGCGGACGCGGCTGTTCC 2085
454 alaAlaGlyValGlyAsnLeuAsnIleSerAspGlyGlyGlySerPheValSer 470
2086 CGCAATGTGCGCAACGTAAGGCGGATTTGGCATTTGACCAATCAGCCCA 2135
471 GlnAsnIleThrPheLeu...GlyAsp..... 478
2136 AGCAGTTTTTGTGTGCGCACCGCATCAAGCCACACCAATCTGTACAGTT 2185
479 .....LysA 480
2186 CGGACGTGACGCGTGTGACAAATTTGTGCAAAAACCATTAACGACGAT 2235
480 lAserGlyIleGlyThrLeuAsnLeuMetAspAlaThrSerSerPheAsp 496
2236 AAGTATTTGCTTCAATGACTAAGACCGACATCAGCGCATATGTCGATCT 2285
497 ThrValGlyIleAsnValGlyAsnPheGly...SerGlyIleValAsnVal 512
2286 TGGCGATCAGCTCATTTAAATCTCAGAGGCTTGCCACACTCAAGCGCA 2335
512 lSerAsnGlyAlaThrLeuAsnSerThrGlyTyrGlyPheIleGlyGly 529
2336 ATCTTAGTGCACAAATGCG.....GATACACGTTAT... 2364
529 snAlaSerGlyLysGlyIleValAsnIleSerThrAspSerLeuThrPasn 545
2365 .....ACAGTCAGCGACCAACGCC..... 2382
546 LeuLysThrSerSerThrAsnAlaGlnLeuLeuGlnValGlyValLeuG 562
2383 ACCCAAAAACGCAACCTTACCTGCTGGGCAATGCCAAGCAACATTTA 2431
562 yThrGlyGlyLeuAsnIleThrThrGlyGlyIleValIleAlaArgAsp 579
2432 ATCAAGCACCAATTAAC.....GGCAACACATGCGCTTGGCG 2469
579 hrGlnIleAlaLeuAsnAspLysSerLysGlyAspValArgValAspGly 595
2470 AATGCTTCA.....TTTAAATCAAGGACACGCGCTTACAAA 2507
596 GlnAsnSerLeuLeuGlyThrPheAsnMetTyrValGlyThrSerGlyPhe 612
2508 CGGAGTGTGACGCTTTCGCGCAACGCTAAGCAACGTA..... 2547
612 rGlyThrLeuThrLeuThrAsnAsnGlyThrLeuAsnValGlyGlyG 629

```


2082 TTCGCCAATGTTGCCAAAGTGAAGCGGATTTGGCAATTTGACCAATGACG 2131
447ThrLeuAlaGlyThrThrLeuHisnslYA 458
2132 CCCAAGCAGTTTGTGTGCGACCGGCAATCAAGCCACACAACTGTACA 2181
458 1a.....IleLeuThr 461
2182 CGTTGCGACTGACGGGTCTGACAAATTTGTGCAAAAAACCATTTACCGA 2231
462 LeuSerGly.....LysThrValAsnAs 469
2232 CGATTAAGTATTCCTTCATTACTAAGACCGACATCAGCGGCAATCTG 2281
469 nAsp.....ThrLeuThrIleArgGluGlyAspAla 480
2282 ATCTTGCGGATCAGGCTCATTTAAATTCACAGGCGCTTGCACAGCTCAC 2331
480 euleuGlnGly.....GlySerLeuThrGlyAsn 489
2332 GGCATCTTAGTGCMAATGGCGATACAGCTTATACAGTCCGCCAAGCG 2381
490 GlySerValGluYsserGlySerGlyThrLeuThrValSer...AsnTh 505
2382 CACCCAAAGCGGCACTTAGCCTCGTGGGCAATGCCCAAGCAACATTTA 2431
505 rThrLeuThrGlnLysAlaValAsnLeuAsnGluGlyThrLeuThrLeuA 522
2432 ATCAAGCCACATTTAAACGCGACACATCGGCTCGGGCAATGCTTCATTT 2481
522 snAspSerThrValThrThrAspAlaIleAlaGlnArgGlyThrAlaLeu 538
2482 AATCTAAGCAGCACCGCGGTACAAAAGCGCAGCTG..... 2517
539 LysLeuThrGlySerThrValLeuAsnGluAlaIleAspProThrAsnVa 555
2518 .ACGCTTTCGCGCAAGCTTAAGCAAAAGCAAGCCATTCGCGCAACG 2566
555 lThrLeuAlaSerGlyAlaThrTrpAsnIleProAspAsnAlaThrValG 572
2567 GTATATGTC.....TCCCTAGCCGATTAAGCAGTATTCATTTGAAGC 2610
572 lnsertValAlaAspAspLeuSerHisAlaGlyGlnIleHisPheThrSer 588
2611 AGCCGCTTAAACGCAAAATCAAGCGCGGCAAGATACGGCATTCACATT 2660
589 ThrArg...ThrGlyLysPhe..... 594
2661 AAAAGACAGCGAATGACGCTGCGTCAAGCGACGGAATTTAGCAATTTAA 2710
595ValProAlaThrLeuLysValLysAsnLeuA 605
2711 ACCCTGACAAGCGCAACATTAACATCAATTCGCGCTATCGCCACAGTGG 2760
605 snGlyValAsnGlnLysThrIleSerLeuArg..... 614
2761 GCAGGGGCGCAACCGGCAAGTGCAGACAGATGCGCGCGCGCTTGGCG 2810
614 614
2811 CGGTCGGCGCGTCCCTATTATTCGGTTACACCGCAACTTCGGTAGAAT 2860
615ValArgProAspMetAlaIleA 622
2861 CCCGTTTCAACAGCTGACGCTGAAGCAAAATTTGAAGCGTACGAGCAACA 2910
622 snAsnAlaAspArgLeuValIleAspGly.....GlyArgAlaThr 635
2911 TTCGCTTATAGTGGAACTCTTGGCTACGACGACCAAAATTTGAAGCT 2960
636 GlyLysThrIleLeuAsnLeuValAlaAlaGlyAsnSerAlaSerGlyLe 652
2961 GCGGAAAGTTCCGAGGCACTTACACTTG...GCGGTCAACAAATACCG 3007
652 uAlaThrSerGlyLysGlyIleGlnValValGluAlaIleAsnGlyAlaT 669
3008 GCAAGCAACCTGCAAGCCTCGACAAATTTGACGAGTAGTGGAAGAAAAAGC 3057
669 hThrGluGluGlyAlaPhe..... 675
3058 AACAAACCGGTGTCGCAAAACCTTATTTACACCTGCAAAACGACACAGT 3107
676ValGlnGlyAsnArgLe 681
3108 CGATGCGCGCGCGTGGCGTTACCAACTC..... 3135
681 uGlnAlaGlyAlaPheAsnTySerLeuAsnArgAspSerAspGlnSerT 698
3136ATCCGCAAAAGCGGAGAGTTCGCGCTGCATTAATCCGTCGAAGAA 3180
698 rPtyrLeuArgSerGluAsnAlaTyrArgAlaGluValProLeuTyrAla 714
3181 CAAAGGCTTTCGCAAAACTCGGCAAGCGCAAGCCAAAAACAGCGCGA 3230
715 SerMetLeuThr..... 718
3231 AAAAGACAGCGGCAAAAGCTTAGCGCGCTGATTTGCGCGCGCGAGTGC 3280
719GlnAlaMetAspTyrArgAlaValAlaGlySerArg.... 731
3281 CCGTCGAAAGACAGAAAGGTTGCCGAACCGCGCGCGAGGCGAGGC... 3327
732SerHisGlnThrGlyAl 737
3328 ...GGGAAATATGTCGCAATTATGCAAGCGGAGGAAGAAAAAGCGGT 3374
738 AsnGlyValAsnAsnSerVal...ArgLeuSerIleGlnGlyLysIle 753
3375 GCAAGGCGATTAAGACACCGCCTTGGCGAAACAGCGCGCAAGCGGAACC 3424
753 uGlnHisAspAsnAsnGlyIleAlaArgGlyAlaThrProGlnSer... 769
3425 GCGCGGCTACACCGCCTTCCCGCGCGCGCGCGCGCGCGGATTTG 3474
770 ...SerIlySerTyrGlyPheValArgLeu..... 778
3475 CCGCAACTGCACACCCCAACCGGACCCCAACCGCGCGCGCTGATCAG 3524
779GluGlyAspLeuMet... 783
3525 CCGTATGCGCAATACGCTTGGAGT.....GAATTTTCGCGCA 3562
784 .ArgThrGluValAlaGlyMetSerValThrAlaGlyValTyrGlyAlaA 800
3563 GCGTCAACAGCGCTTTCGCGTACAGAGCAAGATTAAGCCCGGATTTGCC 3612
800 lAGlnHisSerSerValAspValLysAspAspArgGlySerArgAlaGly 816
3613 GAAGACCGCGCAACCGCCTTGGCAAGCGGACATCCGGGCGCAACCAACA 3662
817 ThrValArgAspAspAlaGlyCysLeuGlyGlyTyrLeuAsnLeuValH 833
3663 CTACCGTTG.....CAAG 3676
833 sThrSerSerGlyLeuThrAlaAspIleValAlaGlnGlyThrArgHis 850
3677 ATTTCCGCGGCTTACCGCAACCAACGACGCGCGCAAACTGCTATGACAG 3726
850 erMetLysAlaSerSerAspAsnAspPheArgAlaArgGlyTyrGly 866
3727 AAAAAGCTCGGACGCGCGCGCTGCGCATCTGTTTCGACCAACCGGAC 3776
867 Trp...LeuGlySerLeuGluThrGlyLeuProPheSer.....IleH 880
3777 CGAAAC.....A 3784

```

880 raspsanleuMetleuGluProGlnleuGlnTyrThrTrpGlnGlyLeuS 897
3785 CCTTCGACGACGCGATCGCAACTCGGACGCGCTTCCGACGGCGCTT 3834
      ::::::::::::::::::::
897 erleuaspaspGlyLysAspAsnAlaGlyTyrValLysPheGlyHisGly 913
3835 TTCGGGACATAGGCGATCGACAGGTTTCATCATGCGC.....AT 3872
      ::::::::::::::::::::
914 SerAlaGlnHisValArgAlaGlyPheArgLeuGlySerHisAsnAspMe 930
3873 CACGGCGGCGCGGCTTTAGCAGC...GGCAGCCTTCAGACGGCATCG 3919
      ::::::::::::::::::::
930 ThrPheGlyGlyThrSerSerArgAlaProLeuArgAspSerAlaL 947
3920 GAGGCAAAATCCCGCGCGCGCTGCTCATTCAGCGATTCGACGACGATC 3969
      ::::::::::::::::::::
947 YshHisValSerGlnLeuProValAsnTrpTrpValGlnPro..... 961
3970 CGGCGCGCTTTCGGCGGATTCGCGCATCGAAACGACATCGGCGACGCG 4019
      ::::::::::::::::::::
962 .....SerValIleArg 965
4020 CTATTTCGTCAAAAGCGGATTACCGCTACGAAACGTCATATTCGCCA 4069
      ::::::::::::::::::::
965 GThPheSerSerArgGlyAspMetArg.....ValGlyThrSer 979
4070 CCCCCGCGCTTCGATTCACCGCTACCGCGCGGCGCATTAAGCAGATTT 4119
      ::::::::::::::::::::
979 hralagly.....SerGlyMet 984
4120 TCATTCGAAACCGCGCAACACATTCGATTCAGCTTATTGAGCTGTC 4169
      ::::::::::::::::::::
985 ThrPheSerProSerGlnAsnGlyThr.....SerLeuAs 996
4170 CTATACGATGCGCTTCGGGCAAAAGTCCGACACGCGTCATATACCGCG 4219
      ::::::::::::::::::::
996 PheGlnAlaGlyLeuGlnAlaArgValArgGluAsnIleThrLeuGly 1013
4220 TATTGGCTCAGGATTTCGGCAAAACCGCGACGTGCGATGGCGGCTGAAC 4269
      ::::::::::::::::::::
1013 alGlnAla..... 1015
4270 GCCGAATCAAGGTTTCAGCGCTGCTCCGACGCGTCCGCGCAAGG 4319
      ::::::::::::::::::::
1016 .....GlyTyrAlaHisSerValSerGlySerAlaGluG 1028
4320 C 4320
1028 Y 1028
seq_name: SwissProt_40:OMP_RICTY
seq_documentation_block:
ID OMP_RICTY STANDARD; PRT; 1645 AA.
AC P96989;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)
DE (rOmp B) (Contains: 120 kDa surface-exposed protein (Surface protein antigen) (120 kDa outer membrane protein ompb); 32 kDa beta peptide).
DE OMPB OR SLP.
GN OMPB OR SLP.
OS Rickettsia typhi.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=785;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WILMINGTON;
RX MEDLINE=94040787; PubMed=8224886;
RA Hahn M.-J., Kim K.-K., Kim I., Chang W.-H.;
RT Cloning and sequence analysis of the gene encoding the crystalline

```

```

RT surface layer protein of Rickettsia typhi.
RL Gene I33:129-133(1993).
RN [2]
RP PARTIAL SEQUENCE.
RC STRAIN=WILMINGTON;
RX MEDLINE=92114896; PubMed=1370573;
RA Ching W.M., Carl M., Dasch G.A.;
RT "Mapping of monoclonal antibody binding sites on CNR fragments of
RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
RT prowazekii."
RL Mol. Immunol. 29:95-105(1992).
RN [3]
RP IDENTIFICATION OF CLEAVAGE SITE.
RX MEDLINE=92104668; PubMed=1729180;
RA Hackstadt T., Messer R., Cleplak W., Peacock M.G.;
RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
RT membrane protein of rickettsiae: identification of an avirulent
RT mutant deficient in processing."
RL Infect. Immun. 60:159-165(1992).
CC -I- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -I- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
CC -I- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC -I- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L04661; ABR48987.1; -.
DR InterPro: IPR003858; rompA_rompB.
DR Pfam: PF02708; rompA_rompB.1.
KW Antigen; S-layer; Transmembrane; Cell wall.
FT CHAIN 1 1353 120 KDA SURFACE-EXPOSED PROTEIN.
FT TRANSMEM 1415 1645 32 KDA BETA PEPTIDE.
FT TRANSMEM 1415 1429 MEMBRANE ANCHOR (POTENTIAL).
FT CONFLICT 657 657 H -> N (IN REF. 2).
FT CONFLICT 842 842 V -> I (IN REF. 2).
FT CONFLICT 1071 1071 G -> A (IN REF. 2).
FT CONFLICT 1306 1306 G -> S (IN REF. 2).
SQ SEQUENCE 1645 AA; 169698 MW; 0CB5641C7EB185EE CRC64;

```

```

alignment_scores:
Quality: 233.00 Length: 1454
Ratio: 0.350 Gaps: 76
Percent Similarity: 45.736 Percent Identity: 20.495

```

alignment_block:

US-09-303-518D-649 x OMPB_RICTY ..

Align seg 1/1 to: OMPB_RICTY from: 1 to: 1645

```

661 GCAATGGCTATTCCTGGCTGTCGGCAATACCTTGACAAAATGG 710
      ::::::::::::::::::::
251 AlaAsnAlaLeuAsnLeuGlnGlyGlyAsnThrIleAsnPheAsnG 267
711 A...TCAGTGTGGCGACGATCAACTAGTAGTGAATAATTAACATA 757
      ::::::::::::::::::::
267 YArgAspGlyThrGlyLysLeuValIleValSerLysAsnGlyAsnAla 284
758 GCCCATATGTTTATTACCAACAGGAGCTATTTGGCGACAGCGGTCA 807
      ::::::::::::::::::::
284 hrGluPheAsn.....ValThrGlySerLeuGlyGlyAsnLeuLys 297
808 CCAATGTTATCTATGAT.....GCCCAAAAGCAAAAGTGTTAAATTA 851
      ::::::::::::::::::::

```

298 GlyValIleGluPheAspThrThrAlaIleValGlyLeuIleAlaAsn 314
852 TGGGGTATTGCAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTCC 901
314 ngly.....GlyAlaIleAsnAlaValIleGlyThrAspAsnGly.... 327
902 AGCTGGTTCGTAAGATTGGTTCTATGATGAACCTTGTGGAGATACC 951
328AlaGly..... 329
952 CATTCAGTATTTACGACACGTCACAAATGGAATACCTTTTAACA 1001
330ArgAlaIleGluPheIleValIleSerValAsn 339
1002 CGATATATATGCG.....ACAGCAAAATCATGCGCAACATGATAC 1042
339 PAsnGlyAsnAlaIleThrIleSerGlyGlnValIleTyrAla..... 352
1043 ACAATTCCTGCTAATGATTAATAAACACGACCCGTCATTTGTTAAT 1092
353LysAspIleValIleGlnSerAlaAsn 361
1093 GTTCTTTATCCGACAGACAGAGAAACCTGTTTATCATGCTGACGCTG 1142
362 AlaGlyGlyGlnValIleThrPheGlnHisLeuValAspValGlyLeuGly 378
1143 TGTCAACAGTTATCGA.....CCAGACTGAAATATGAGAAAAATA 1183
378 YLysThrAsnPheLysThrAlaAspSerLysValIleIleThrGluAsn 395
1184 TTTCCTTATATGACGAGAAAGGCGCAATG..... 1215
395 LAspPheGlySerThrAspPheGlyAsnLeuAlaValGlnIleVal 411
1216AFACTTACCACCAACATC.....AATCA 1238
412 ProAsnAsnLysIleLeuThrGlyAsnPheIleGlyAspAlaLysAsn 428
1239 AGGT.....GCTGAGGATTAATATTCAGAGAGATTTACGCTGCC 1282
428 nGlyAsnThrAlaGlyValIleThrPheAsnAlaAsnGlyThrLeuVal 445
1283 CTGAAATTAACGAA..... 1296
445 eGlyAsnThrAspProAsnIleValIleThrAsnIleLysAlaIleGlu 461
1297ACTTGCAAGCGCGCGCTCATATGATGAA.. 1329
462 ValGluGlyAlaGlyIleValIleGlnLeuSerGlyIleHisGlyAlaGlu 478
1330GACAGTACCGTTA 1342
478 uArgLeuGlyAsnAlaGlySerIlePheLysLeuAlaAspGlyThrVal. 494
1343 CTTGGAATTAACGGCGCGCAACGAC.....CGCGCTGCC 1380
495IleAsnGlyProValAsnGlnAsnProLeuValAsnAspAsn 508
1381 AAAATCGGCAAGGACGCTGACGCTTCAAGCCAAAGGAGGAAACCAAG 1400
509 AlaLeuAlaIleGlySerIleGlnLeuAspLysSerAlaIleIleThrGly 525
1431 CTCGATCAAGCGTGGCGACGATGCTATTTGGATCAGCAG..... 1473
525 yAspIle.....GlyAsnGlyAlaValAlaAsnAlaIleGlnAspIleThr 540
1474GCAGACGATTAAGGCAAAACCAACGCTTACGAAATCGCGCTG 1518
540 hLeuAlaIleAsnAspAla.....SerLysIleLeuThr 550
1519 CTCAGCGGACGAGGTACGCTGCAATGATGCCGATATCATGTCACACC 1568
551 LeuSerGlyAlaAsnIleIleGlyAlaAsnAlaGlyGly..... 563

1569 CGACAACATCTATTTCGGCTTCCGGCGGACGTTTGATTTA..... 1611
564AlaIleHisPheGlnAlaIleAsnGlyGlyThrIleGlnLeuThrSer 579
1612ACGGGCAATTCGCTTTCGTTCCACCGTATTCAAATACCGATGA 1656
579 hArgAsnAsnIleLeuValAspPheAspLeuAspValIleThrThrAspGln 595
1657 GGGGCGATGATTTGCAACACCAATCAACACAAAGATTCACGCTTACAT 1706
596 ThrGlyAlaValAspAlaSerSerLeuThrAspAsnGlnThrLeuThrIle 612
1707 TACAGCAATTAAGATATTGCTACACCGGCAATTAACACAGCTTGATA 1756
612 eAsnGlySer.....IleGlyThrIleGlyAlaAsn..... 622
1757 GCAAAAAAGAAATTGCTTACACCGCTTGCTTGGCGAAGAAATACGACC 1806
623Thr 623
1807 AAAACGAAACGGCGGCTCAACCTGTTTACACCGCGCGCAAGACCG 1856
624 LysThrLeuGlyArgPheAsnVal.....GlySerSerLys 635
1857 CACCGCTGCTTCCGCGGCAACAAATTTAAACGGCAACATCAGCAAA 1906
635 sThrIleLeuAsnAlaGlyAspValAlaIleAsnGlnLeuValMetGlu 652
1907 CAACGGCAACATGTTTTCAGCGGCGAGAACCAACACCGGACGCTTAC... 1953
652 snAspGlySerValHisLeu.....ThrHisAsnThrTyrLeu 664
1954AATCATTTAAACGACCATTTGGTCCGCAAAAGAGG 1988
665 IleThrLysThrIleAsnAlaIleAsn..... 673
1989 CATTCCTCGCGGCGGCAATGCTGGGAGCAACGATCGATCAACGCC...A 2035
674GlnGlyLysIleIleValAlaIleAsnProIleAsnThrAsp 688
2036 CATTTAAACGGCAAACTTCAAAATTAAGGCGGACGCGCTGCTTCC 2085
688 hAlaLeuAlaAspGlyThrAsnLeuGlySerAlaGlnSerProLeuSer 704
2086 CGC.....AATGCTTCCAAAGTGAAGCGGAT.....TGCAATTT 2120
705 AsnIleHisPheAlaThrLysAlaIleAsnGlyAspSerIleLeuHisIle 721
2121 GAGCAATCAGCGGCAAGGATTTTGTGTCGACACCGCATCAAGGCA 2170
721 eGlyLys.....GlyValAsnLeuTyrAlaAsnAsnIle 732
2171 CAATCGTACAGCTTCGAC.....TGACGCGCTG 2202
732 LeThrThrThrAspAlaAsnValGlySerLeuHisPheArgSerGly 748
2203 ACAAAATGCTGCAAAAAACCATTAACGAGATTAAGATGCTGTTAT 2252
749 ThrSerIleValSerGlyThrValGlyGlyGlnLeuLys..... 763
2253 GACTAAGACCGACATCAGCGGCAATGCTGATTTGCGGATCAGCTCAT 2302
764LeuAsnAsnLeuIleLeuAspAsnGlyThrThr 775
2303 TAAATCTCAGAGGCTTCCCACTCAACGCAATCTTATGCAATGCG 2352
775 aLysPheLeuGlyAspIleThrPheAsnGlyGlyThrLysIleGlnGly 791
2353 GATCACAGCTTATACAGTACAGCACACAC..... 2379
792 LysSerIleLeuGlnIleSerSerAsnTyrIleThrAspHisIleGlu 808

2380 .GCCACCAAAAGGCAACCTTAGCCTGGGCAATGCCCAACACAT 2428
|||
808 rAlAspAsnThrGlyThrLeuGluPheValAsnThrAspProLeuThr. 824
2429 TTAATCAGCCACATTAAAGGCAAC..... 2454
|||
825ValThrLeuAsnLysGlnGlyAlaTyrPheGlyValLeuLys 838
2455ACATCGGCTCGGGCAATGCTCATTTAATCTAAGGCA.. 2493
|||
839 GlnValMetValSerGlyProGlyAsnIleAlaPheAsnGluIleGlyAs 855
2494CAGCCGTA.....CAAAAGGCA 2512
|||
855 nGlyValAlaHisAlaIleAlaValAspSerIleSerPheGluAsnAla 872
2513 GTCTG.....ACGCTTTCGGC..... 2529
|||
872 erLeuGlyAlaSerLeuPheLeuSerGlyThrProLeuAspValLeu 888
2530 AAGCTTAAGCAACGTAACCATTCGCGCATCAACGTAATGTCTCCT 2579
|||
889 ThrIleYSerThrValGlyAsnGlyThrValAsp...AsnPheAsnAl 904
2580 AGCCGATAGGCAGATATTCATTGTAAGACGCCCTTACCGGACAA 2629
|||
904 aProIleLeuValValSerGlyIleAspSerMetIleAsnAsnGlyGlnV 921
2630 TCACCGGCGGCAAGGATACGATACGATTA.....AAGACAGC 2670
|||
921 aIleGlyAspGlnLysAsnIleIleAlaLeuSerLeuGlySerAspAsn 937
2671 GAATGAGCGCTCGCTACGACGACGAAATTAGCCAAATTA.....AA 2711
|||
938 SerIleThrValAsnSerAsnThrLeuTyrAlaGlyIleArgThrThr 954
2712 CTTGACAAAGCCACCATTAACATTCGCTATCGCCACGAT.... 2757
|||
954 sThrAsnGlnGlyThrValThrLeuSerGlyGlyIleProAsnAsnPro 971
2758GCGCAGGGGCGCAACCGGACGCTGCGACAGATGCGCG 2796
|||
971 LyThrIleTyrGlyLeuGlyLeuGluAsnGlyAspPro..... 983
2797 CGCCCGCTTCGCGCGCTTCGCGCGTTCCTATTATCGTTACACGCC 2846
|||
984LysLeuLysGlnValThrPheTh 991
2847 AACTTCGTAGATCCGTTTCAACACGCTG..... 2877
|||
991 rThrAsp.....TyrAsnAsnLeuGlySerIleIleAlaThrA 1004
2878ACGCTAAC..... 2886
|||
1004 sNValThrIleAsnAspAspValThrLeuThrThrGlyGlyIleAlaGly 1020
2887GCAATTTGAACGGTCAAGGAC 2909
|||
1021 ThrAspPheAspGlyLysIleThrLeuGlySerIleAsnGlyAsnAlaAs 1037
2910 ATTCCGCTTATGTCGAACCTTTGCGCTACCGCAGCACAATTTGAAGC 2959
|||
1037 nValLysPheValAspArgThrPheSerHisProThrSerMetIleValS 1054
2960 TGCGGGAAGTTCGCAAGCAGCTTACAC.....TTGCGGCTC 2997
|||
1054 erThrLysAlaAsnGlnGlyThrValThrTyrLeuGlyAsnAlaLeuVal 1070
2998 AACATATACCGCAACGAA.....CCTGCAAGCTTCGAACATTTACGCT 3041
|||
1071 GlyAsnIleGlySerSerAspIleProValAlaSerValArgPheThr.. 1086
3042 AGTGAAGAGAAAGACAAACCGCTGTCGAAACCTTAATTTCACCC 3091
|||||
1087GlyAsnAspSer.....GlyValGlyLeuGlnGlyAsnI 1098
3092 TGCAAAAGCAACACGTGATCCGCGCGCTTACCAATCATTCGCG 3141
|||||
1098 lHisSerGlnAsnIleAspPheGlyThrTyrAsnLeuThrIleLeuAsn 1114
3142 AAGAGC.....GCCGATTTCG 3158
|||||
1115 SerAspValIleLeuGlyGlyThrThrAlaIleAsnGlyGluIleAs 1131
3159 CCTGCATTAATCCGTCGAAGAACAGACCTTCCGACAAATTCGCGCAG 3208
|||
1131 pLeu.....LeuThrAsnAsnLeu..... 1137
3209 CAGAACCAAAACAGCGCGCAAAAGAACAGCGCAACGCTTAGCGC 3258
1137 1137
3259 CTGATTGCGCGCGCGCGATTCGCGAAGACAGAAAGCTTTCGCGA 3308
|||||
1138 IlePheAlaAsnGlyThrSerThrTrpGlyAsnAsnThrSerLeuSerTh 1154
3309 ACCGCGCCGCGCAGCGCGCGGAAATGTGCGC...ATTATGACGCGC 3355
|||
1154 rThrLeuAsnValSerAsnGly...AsnValGlyGlnIleValIleAlaG 1170
3356 AGGAGAGAAAGAAACGGTGACGCGGATTAAGACACGCGCTTGCGAAA 3405
|||
1170 LuGlyAlaGlnValAsnAlaThrThrGlyThrThrThrIleLysIle 1186
3406 CAGCGCAGACCGCAAAACCGCG...CGGCTACACCGCGCTTCCCGCGC 3452
|||
1187 GlnAspAsnAlaAsnAlaAsnPheSerGlyThrGlnThrTyrThrLeuI 1203
3453 CCGCGCGCGCGCGGATTTGCCGCACTGCAACCCCAACCGCAGCCCC 3502
|||
1203 eGlnGlyGlyAlaArgPheAsnGlyThrLeu..... 1213
3503 AACCGCAGCGGACCTGATACCGCTTATGCAATAGCGTTTGATGA 3552
1214GlyAlaProAsn 1217
3553 TTTTCGCGACGCTCACACAGGTTTTCGCGCTACAGACGAAATTAGCG 3602
|||||
1218 PheAspValThrGlyAsnAsnIlePhe...ValLysTyrIleLeu..... 1231
3603 GGTATTTCGCAAGACCGCGCAACGCCGTTTGACAAAGCGGATCCGG 3652
1232IleArg 1234
3653 ACACCAAACTACCGTTTCGCAAGATTTCGCGCTACCGCAACAAAC 3702
|||||
1234 sPAla.....AsnGlnAspTyr..... 1239
3703 GACTGCGCCAAATCGTATGCAAGAAACCTCGCGACGCGCGCTCG 3752
1239 1239
3753 CATCTCTTTTTCGCAACCGGACCGAAACCTTCGACGACGCGCATCG 3802
|||||
1240 .ValLeuThrArgThrAsnAspValLeuAsnValValThrThrAlaValG 1256
3803 GCACTCGCAGCGCTTCGCCACGCGCGCG.....TTTTCGGCAATAC 3846
|||||
1256 LysAsnSerAla...IleAlaAsn..AlaProGlyValHisGlnAsnIleAl 1271
3847 GGCATGACAGGTTTACATCGGCAATCAGCGCGCGGCGGTTTATAGAG 3896
|||||
1271 aIleCysLeuGlnUserThrAspThrAlaAlaTyrAsnMetLeuLeuA 1288
3897 CG.....GCAGCTTTCAGACGCGCATTCGAGGCAAAA 3928
|||

```

1288 lalysaspSerSeraspValAlaThrPheIleGlyAlaIleAlaThrAsp 1304
3929 TCCGCCCGCGCTGCTGATTCAGCATTCAGCAGCAGTACCGCCCGCT 3978
      :::::||||| :::::||||| :::::|||||
1305 ThrGlyAlaIleAlaValAla...ThrValAsnLeuAsnAspThrGlnLys... 1319
3979 TTGGCGGATTGGCGATCGACCGCAGCATTCGCGCAGCGCTATTTC... 4026
      :::::||||| :::::||||| :::::|||||
1320 ThrGlnAspLeuLeuGlyAsnArg...LeuGlyAlaLeuAlaGlyTyrLeuSe 1335
4027 .....
1335 rAsnSerGluThrAlaAspValGlyGlySerGluThrGlyAlaValSerS 1352
4034 AAGCGGATTACCGCTACGAAAAAGTC..... 4059
1352 eGgIlyAspGluAlaIleAspGlnValSerTyrGlyValTyrPalLysPro 1368
4060 .....AATATGCCACCCCGC.....GCCCTTGA...TTCAA 4088
1369 PheTyrAsnIleAlaGluGlnAspLysGlyGlyLeuAlaGlyTyrLys 1385
4089 CCGCTACCGCGCGGATTAAGCAGATTATTCATTCAAACCGCGCGCAAC 4138
      :::::||||| :::::||||| :::::|||||
1385 sAlaLysThrAlaGlyValAlaValGlyLeuAspThrLeuAlaAsnSPa 1402
4139 ACATTTCATCAGCGCTTATTGAGCTGTCTATACCATGCGCGCTTCG 4188
      :::::||||| :::::|||||
1402 snLeuMetIleGlyAlaAlaIleGlyTyrThrLysThrAsp..... 1415
4189 GCGAAAGTCCGACACGCGTCAATACGCGTATGTCGATGCGATTCGG 4238
1416 .....IleLysHisGlnAspTyrLys 1422
4239 CAAACCGCGATGCGGATGCGGCTAAACGCGGAATCAAGCTTTCA 4288
      :::::||||| :::::|||||
1422 sLys.....GlyAspLysThrAspLysLysLysLys 1433
4289 CGGTGCTCCACGCTGCGCGCGCGCAAGCGCGCAACTGAGACGCA 4338
      :::::||||| :::::|||||
1433 eRPeSerLeuTyrGlyAlaGlnGlnLeuValLysAsnPhaIleGln 1449
4339 CACAGCGCG 4347
      :::::|||||
1450 GlySerAla 1452

```

seq_name: SwissProt_40:AMYH_YEAST

seq_documentation_block:

```

ID AMYH_YEAST STANDARD; PRT; 1367 AA.
AC P08640; P08068;
DI 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glucoamylase SI/S2 precursor (EC 3.2.1.3) (Gluca 1,4-alpha-
DE glucosidase) (1,4-alpha-D-glucan glucosylhydrolase).
GN STR1 OR STR2 OR MALS OR YIR019C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5286C / AB972;
RA Bartell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule S., O'Neill C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skellon J., Smith V.,
RA Welsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RP [2]
SEQUENCE OF 1-242 AND 762-1331 FROM N.A.

```

```

RX MEDLINE=871944600; PubMed=3106330.
RA Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
RT STR1".
RL J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN=SPX101-1C;
RX MEDLINE=89031230; PubMed=3141213;
RA Pardo J.M., Ianez E., Zalacain M., Clavos M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the STR2 and SGA genes
RT from Saccharomyces cerevisiae".
RL FEBS Lett. 239:179-184(1988).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -1- SIMILARITY: TO S.POMBE SPC215.13.
CC -1- SIMILARITY: SOME, TO S.POMBE SPC285.13C.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
DR EMBL: Z38061; CAA8616.1; -
DR EMBL: M16164; AAA35014.1; -
DR EMBL: M16165; AAA35015.1; -
DR EMBL: X13857; CAA32069.1; -
DR PIR: B26877; B26877.
DR PIR: A26877; A26877.
DR SCD: S46478; S48478.
DR SGD: S0001458; MOC1.
KW Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal; Multigene family.
FT SIGNAL 1 21
FT CHAIN 22 1367
FT DOMAIN 210 1367
FT CARBOHYD 817 817
FT CARBOHYD 874 874
SQ SEQUENCE 1367 AA; 136110 MW; 91C00F2DBD61AA9D CAC64;

```

alignment_scores:

```

Quality: 232.50 Length: 1098
Ratio: 0.450 Gaps: 46
Percent Similarity: 47.086 Percent Identity: 20.036

```

alignment_block:

US-09-303-518D-649 x AMYH_YEAST

Align seg 1/1 to: AMYH_YEAST from: 1 to: 1367

```

1186 TCCCTTATTGACGAGAAAGCGATTTGATCTTACCGACCAATCA 1235
      :::::||||| :::::||||| :::::|||||
25 AlaLeuValAlaProArgLysSerSerGluGlyThrSerGlyAsnSerIleVal 41
1236 TCAAGCTGCTGAGGATTTATTTTCAAGAGAGATTTACGCTCGCGT 1285
      :::::||||| :::::|||||
41 LAsnGlycysProAsnLeu.....AspPheAsnTyrPheIleMetA 54
1286 AAAATACGAAACTTGGCAAGCGCGCGCTGATATCAATGAGACAGT 1335
      :::::||||| :::::|||||
54 sPngInGlnAsnIleMetCln.....TyrThrLeuAspValThr 66
1336 ACCGTTCCTTGAGAAAGTAAAGCGGCTGGCAAGACCGCGCTGCAAAAT 1385
      :::::||||| :::::|||||
67 SerValSerTrp.....ValGlnAspAsnThrTyrGlnIle 78
1386 CCGCAAGGACGCTGACGCTTCAAGCAAGGCGCAAGACGCTGCA 1435
      :::::||||| :::::|||||

```



```
78 e.....ThrIleHisValIysGlyLys...GluAsn.....87
1436 TCAGCGTCGGCAGCGTACAGCTATTTGGATCAGCAGCGACGATATAA 1485
88 .....LeuAspLeu 90
1486 GGCATAAAACACAGCCTTAGTAATCGCTTGTCAGCGCAGGCGTAC 1535
91 LysTyrLeuTyrPseSerLeuLysIleIleGlyValThrGlyProLysGlyTh 1570
1536 GGTGCACATGTAATGCCGATTAATCACTTCAACC.....CCG 1570
107 rValGlnLeuTyrGlyTyrAsnGlu.AsnThrTyrLeuIleAspAsnPro 123
1571 ACNAACTGATTCGGCTTC.....GC 1593
124 ThrAspPheThrAlaThrPheGluValTyrAlaThrGlnAspValAsnse 140
1594 GCGCGACGCTTGGATTAAACGGGCATTGCTTGGTTCCACCGCTATTCA 1643
140 rCysGlnValTyr..MetProAsnPheGlnIleGlnPheGluTyrLeuG1 156
1644 AATTACCGATGAAGGGCGCATGTTGTCAACCAACATCAAGCAAGAAT 1693
156 nGlySerAlaAla.....GlnTyrAlaSerS 165
1694 CCACCGTTACATTACAGCAGCAATTAAGATTGTCTACAAACGGCAATAAC 1743
165 erThrPglntPrGlyThrThrSerPheAspLeuSerThrGlyCysAsnAsn 181
1744 AACAGCTTGATAGCAAAAAAGAAATGGCTACACGGGTGGTGGGCA 1793
182 TyrAspAsnGlnGlyHisSerGlnThrAspPheProGlyPheTyrTrpAs 198
1794 GAAAGAT.....ACGACCA 1807
198 nIleAspCysAspAsnAsnGlyGlyThrLysSerSerThrThrThrS 215
1808 AACGAAAGGGGGCTCAACCTGTTTACAGCCCGCCGACAGAAAGCCGC 1857
215 erSerThrSerGlnSerSerThrThrThrSerThrSerThrSerThrS 231
1858 ACCGCGCTTTCGGCGGACAAATTTAAACGGCAACATCAACGCAAAC 1907
231 rThrThrThrSerThrSerThrSerThrThrThrThrSerThrS 248
1908 AACGCGCAACTGTTTTCAGCGGACACACAAACCCGACCGCTACAAATC 1957
248 erGlnSerSerThrSerSerSerThrThrAlaProAlaThrProThr... 263
1958 AATTAAAGCAGCATTTGGTCGCAAAAAGAGGCAATTCCTCGCGGGAAATC 2007
264 .....ThrThrSerCysThrLysGluLys.....ProTh 273
2008 GTGTGGGCAACAGACTGATCAACCGCACATTTAAACGGGAAAACTTCCA 2057
273 rProProThrThrThrSerCysThrLysGlu..LysPro..... 285
2058 AATTAAAGCGGACAGAGGGGTGTTCCGCAATGTTCGCAAACTGAAG 2107
285 ..... 285
2108 GCGATTGGCATTTGAGCAATCAGCCCAAGCATTTTGGTGTGCGACCG 2157
286 .....ThProPro 288
2158 CATCAAGCGACACATCTGTACAGTTCGACGACGAGCGGTGACAA 2207
289 HisHisAspThrThrProCysThrLysLysLysThrThrThrSerLysTh 305
2208 TTGTGTGCAAAAAACCATTAACGACAGATTAAGTGAATGCTTCATGACTA 2257
305 rCysThrLysLysThrThrThrProValProThrProSerSerSerThr 322
2258 AGACCGACATCAGCGGCATGTCGATCTTCCGATCAGCGCTATTAAAT 2307
322 hrGlnSerSerSerAlaProValPro..... 330
2308 CTCACAGGCGTTGCCACATCAACGGCAATCTTAGTGAATGGCGATAC 2357
330 ..... 330
2358 ACGTTATACAGTCAAGCCACAGCCACCAAAAGCGCAACCTTAGCCTCG 2407
331 .....ThrProSerSerSerThrThrGlnSer..... 339
2408 TGGGCAATGCCCAAGCAACATTAATCAAGCCACATTTAAAGCCCAACACA 2457
340 .....SerSerAlaProValThrSerSerThr 348
2458 TCGGCTTCGGGCATGCTTCAATTATATCAAGCAGCAGCCGCTACAA 2507
349 ThrGlnSerSerSerAla.....ProValProTh 358
2508 CGGACGTCTGACGCTTTCGGCAACGCTAAGCAAGTAAGCCATTCCG 2557
358 rProSerSerSerThrThrGlnSerSerSerSerAlaProValThrSerS 375
2558 CACTCAACGCGTAATGTCTCCCTAGCCGATTAAGCAGATTTCCATTGAA 2607
375 hrThrGlnSerSer..... 379
2608 AGACGCGCGTTTACCGGACAAATCAGCGCGGCAAGATTAAGGATTAACA 2657
380 SerAlaProValThrSerSerThrThrGlnSerSerSerSerAlaProValPr 396
2658 CTTAAGAACAGCAGGAATGACGCTGCCCTCAGACGACGCAATTAAGCAATT 2707
396 oThrProSerSerSerThrThrGlnSerSerSerSerAlaProValThrSers 413
2708 TAAACCTGACAGCGCCACCATTAACCTCAATTCGCGCTATCCGACGAT 2757
413 erThrThrGlnSerSerSerSerAlaProValThrSerSerThrThrGlnSer 429
2758 GCGGACAGGGCGGCAAAACCGGACGTGACAGATGCGCGCGCGC.... 2802
430 SerSerAlaProValThrSerSerThrThrGlnSerSerSerSerAlaProva 446
2803 .CGTTCGCGCGCTTCGGCGCGTCCCTATTATCCGTTAACCGCGCAACTT 2851
446 lThrSerSerThrThrGlnSerSerSerSerAlaProValProThrProSers 463
2852 CGGTAGAAATCCGTTTTCACAGCAGCTGACGTTAAACGGCAATTTGAACGT 2901
463 erSerThrThrGlnSerSerSerSerAlaProVal.Thr..... 474
2902 CAGGCAACATTCGCTTATGTGAGACCTCTTGGGCTACCGCAGCAGCAA 2951
475 .....SerSerThrThrGlnSerSerSerSerAlaProValProThr. 487
2952 ATTGAAGCTGCGGAAAGTTCCGAAAGCAGCTTACACCTTGCGGCTACACA 3001
488 .....ProSerSerSerThrThrGlnSerSerSer 497
3002 ATACCGGACAGCAACCTGCAAGCCTCGAACATTTGAGGATGAGAGGA 3051
498 AlaProValThrSer..... 502
3052 AAGACACAAACCGCTGTCGAAACCTTAATTTACCCCTGCAAAAGCA 3101
503 .SerThrThrGlnSerSerSerSerAlaProValProThrProSerSerS 519
3102 ACACG...TCGATGCCGGCGCGCTGCGTTACCACTCATTCGCAAAAGC 3148
519 hrThrGlnSerSerSerAlaProAlaProThrProSerSerSerThrThr 535
```


1500 CTTTACGAATCGGCTTGTGTCAGCGGCGAGGTACGGTG...CACTGA 1546
1534 aleuGlnHisIleThrLeuAlaAsnAspAlaSerLysIleLeuAlaLeu 551
1547 ATGCCGATATGATGTTCAACCCCGAC.....AACTCTATTTCGGC 1587
551 sp61yAlaAsnIleIleGlyAlaAsnValIglYAlaIleHisPheGln 567
1588 TTTCCGGGCGACGTTTGGATTTAACGGCGCATTCGCTTCGTTCCAGC 1637
568 AlaAsnIglYIle.....Ly 574
1638 TATTCAAAATACCGATGAGGCGGATGTTGTCACCAACATCAAGACA 1687
574 sleuthrAsnThrGlnAsnIleVal.....AsnPheAsp 588
1688 AAGAAATCCACCGTTACCTTACAGGCAATAAGATATGCTACACCGCG 1737
588 euAspIleThrAspLysThrIglYAlaValAspAlaSerLeuThr 604
1738 AATACACACGCTTGATAGCAAAAGAAATGCGCTACACCGTTGTT 1787
605 AsnAsnGlnThrLeu.....ThrIleAsnIglYSerIle 615
1788 TGGCGAG...AAGATACGACCAAAACGAGCGGCGCTCAACCTGTT 1834
615 eglYThrValAlaAsnThrLysThrLeuAlaGlnLeuAsnIle... 630
1835 ACCGACCGCGGAGAGACCGGACCGCTGCTTCCGCGGACAAAT 1884
631GlySerSerLysThrIleLeuAsnAlaGlyAspAla 643
1885 TTAAACGCGCAACATCAACGCAAAACGCAACGCTTTTTCAGCGGCG 1934
644 IleAsnIleValIleGlnAsnAsnIglYSerVal..... 655
1935 ACCACACGCGACGCTACATGATTTAAACGACGCTTGTGCGCAAAA 1983
656GlnLeuAsnHis...AsnThrLysLeuIleThrLys 667
1984GAGGCAATTCCTCGCGGGAATGCTGAGACACGACTGAG 2028
667 hrIleAsnAlaIleAsnGlnIglYIleValAlaIleAspProLeu 683
2029 AACCGCACATTT.....AACCGCAAAA 2051
684 AsnThrAsnThrThrLeuAlaAspIleThrAsnLeuGlySerAlaGln 700
2052 C.....TTCCAAATTAAGCGCGACGCGGCGGTT 2083
700 hrProLeuSerThrIleHisPheAlaThrLysAlaIleAsnAlaAspSerI 717
2084 CCGCAATGTGSCAAAGGAAAGCGGATTTGAGCAATCAGCGC 2133
717 leuAsnValIglYs..... 722
2134 CAAGCAATTTTGTGTGTCAGCGCATCAAGCCACAAATCTGTACAG 2183
722 722
2184 TTCGACGTCGACGCGTTCGACAATGTGTGCAAAAAACCATTCAGCAG 2233
723GlyValAsnLeuIglYAlaAsnAsnIleThrThrAsn 725
2234 ATAAAGTATGCTTCATG.....ACTAGACCGGACATC 2268
725 spAlaAsnValIglYSerLeuHisPheArgSerIglYThrSerIleVal 751
2269 AGCGGC.....AATGTCAGTCT 2285
752 SerGlyThrValIglYIglYngIglYHisLysLeuAsnAsnLeuIleLe 768

2286 TCCGATCAGCGCTCATTTAAATCTCACAGGCGTTCACACTACGCGCA 2335
768 uAspAsnGlnThrThrValLysPheLeuGlyAspThrIlePheAsnIgl 785
2336 ATCTTAGTGCAAATGCG.....GATACACGCTTATACA 2367
785 IglYThrLysIleGlnIglYsSerIleLeuGlnIleSerAsnAsnIglYThr 801
2368 GTCACCGAC.....AACGCCCAAAACGCGCAACTTACGCTGCG 2411
802 ThrAspHisValIglYSerAlaAspAsnThrIglYThrLeuGlnPheValAs 818
2412 CAATGCCCAACCAACATTTATCAAGCCACATTAACGCGAAC..... 2454
818 nThrAspProIleThr.....ValThrLeuAsnLysGlnIglYAla 832
2455ACATGCGCTTCGCGCAATGCTTCA 2478
832 yrPheGlyValLeuLysGlnValIleIleSerGlyProGlyAsnIleVal 848
2479 TTTAAT.....CTAAGCA 2492
849 PheAsnGlnIleGlyAsnValIglYIleValHisGlyIleAlaIleAsnSe 865
2493 CCACGCGGTACAAACGCGACGCTGACGCTTTC..... 2526
865 rIleSerPheGlnAsnAlaSerLeuGlyThrSerLeuPheLeuProSerG 882
2527GCCAAGCTAAGCAACGTAAGCAATTCGCGCA 2559
882 IglYThrProLeuAspValLeuThrIleLysSerThrValIglYAsnIglYThr 898
2560 CTCACGCTAATGTCTCCCTACGCGATTAAGCGGATTTTCATTTGAAG 2609
899 ValAsp...AsnPheAsnAlaProIleValIleValSerGlyIleAspSe 914
2610 CACCGCGTTAACCAGCAATCAACGCGCGCGGCAAGATACGCTATACAT 2659
914 rMetIleAsnAsnIglYIleIleGlyAspLysLysAsnIleIleAla 931
2660 TA.....AAGCACGCGAATGAGCGCTGCGCGGCGCAAAATTA 2700
931 euSerLeuGlySerAspAsnSerIleThrValAsnAlaAsnThrLeuYr 947
2701 GCGAATTTAACTT.....GACAAACGACCATTAACGTAATTC 2741
948 SerGlyIleArgThrThrLysAsnAsnGlnIglYThrValThrLeuSerG 964
2742 CGCCTATCGCCACGAT.....GCGGACGCGCGGCAACCG 2776
964 yglMetProAsnAsnProGlyThrIleTyrGlyLeuGlyLeuGlnG 981
2777 GCAGTGCACACAGATCGCGCGCGCGCTTCGCGCGGCTTC 2826
981 LysPro.....Lys 984
2827 CTATATCGCTTACCGCGCACTTCGCTAGATCCCGT..... 2865
985 LeuLysGlnValThrPheThrThrAspLysAsnLeuGlySerIleI 1001
2866 TTCACACGCGTGAAGGTAAC..... 2886
1001 eAlaAsnAsnValThrIleAsnAspTyrValThrLeuThrGlyIglY 1018
2887GCCAATTAAGCGGT 2901
1018 leaIglYThrAspPheAspAlaLysIleThrLeuGlySerValAsnGly 1034
2902 CAGGACATTCGCTTATGTCGGAATCTTCGCTACGCGACGACGA 2951
1035 AsnAlaAsnValArgPheValAspSerThrPheSerAspProArgSer 1051
2952 ATTGAAGCTGCGGAAGTTCGGAAGGCACTTACAC.....T 2989

[illegible]

681	SerLeuAlaIleThrGlyHisAlaLysGluValGluAspLysGlnTyrSer	697
1699GTATCCATTACAGCAATATAAGATATTGCTACA....	1731
697	IaIaGlyPheHisIleThrHisThrLysAsnThrSerThrGlnTyr	714
1732ACCGCAATTAACAACAGCTTGATGCTAGC	1758
714	hrcGluGlnAlaAsnSerThrIleSerGlyAlaAsnValAspLeuGlnAla	730
1759	AAAAAGAAATTCCTCCACAACGGTGTGGTGGCAGAGAAAGATACGCCAA	1808
731	AsnLysAspValThrPheIleAsnGly.....SerAspLeuLysThrThrAl	745
1809	AACGACAGGGCGGCTC.....AACCTGTCTTAC.....	1836
745	agLysAlaSerIleThrGlyAspAsnValAlaPheValSerThrGluA	762
1837	..CAGCCCGCGCGAGAACGCCACCTCGTGTCTTCCGCGGAACAAT	1884
762	snLysLysGlnThrAspAsnThrAspThrThrIleSerGlyGlyPheSer	778
1885	TTAAACGGCAACATCACGCCAAACAACGGCAACTGTTTTCAGCGGCAG	1934
779	TyrThrGlyGlyValAspLysValGlySerLysAlaAspPheGln.Tyra	795
1935	ACCAACACCGCAGCCGTACATCATTTAAACGACCATTTGGTCGCAAAAG	1984
795	sPLysGlnHisIleGlnThrGluVal.....ThrLysAsnArgLysSer	809
1985	AGGGCATTCCTCGCGGGGAAATCGTGTGGACACACGACTGGATCAACGC	2034
810	GlnThrGluValAlaGlyAspLeuThrIleThrAlaAsnLysAspLeuLe	826
2035	ACATTTAAAGCGGAACACTCCAAATTAAGCGCGGACAGCGGTGGTTTC	2084
826	hHis..GluGlyAlaSerHisHisValGluGlyArgTyrGlnGlnSerG	842
2085	CCGCAATGTGGCAAAAGTAAGAAAGCGATGGCATTTGACCAATCACGGCC	2134
842	ylGluAsnIleGln.....HisLeu.....	848
2135	AAGCACTTTTGGTGTGCGACCGCATCAAGACACACATCTGTACACGT	2184
849AlaValAsnAspSerGlnThrSerLysThrAsp	859
2185	TCG.....GACTGACGGGGTC	2201
860	SerLeuAsnValGlyIleAspValGlyValAsnLeuAspLysSerIleVal	876
2202	GACAAATTCGTGGAAAAACCAATTCGACGACGATTAAGATGATTCCTCAT	2251
876	IThrLysProValLysLysAlaIleGluAspGly.....ValAsnTr	890
2252	TGACTAAG.....ACGCACATCAGCGGCATGTGCATCTTGCC	2289
890	hThrLysProGlyAsnAsnThrAspLeuThrLysValThrAlaArg	906
2290	GATCAC..GCTCATTTA..AACTCCAGAGCGCTTCGACACTCAAC..	2311
907	AspAlaIleAlaAsnLeuAlaAsnLeuSerLysLeuGlnThrProAsnVal	923
2332GSCAATCTTAAGTCAAAATGGCG	2353
923	IcLysAlaGluValGlyIleLysGlyGlySerGlnGlnSerGlnThrA	940
2354	ATACACGT...TATCACATCAGCCCAACAGCC.....	2382
940	spSerGlnAlaValSerThrSerIleAsnAlaGlyLysAlaAspLysLe	956
2383ACCAAAA	2399
957	SerAsnAsnLysLeuHisAspGlnTyrHisTyrGlnSerThrGlnGln	973

2391 CGGCACCTTACGCTGCGCAATGCC..... 2418
973 ugiL.....IleSerLeuThrAlaSerHisThrSerGluAlaThrLeuA 989
2419CAAGCAATTTAATCAAGCCATTAAGCGCAACGCAACATCG 2460
989 sprLysHisGlnThrThrPheHisGlnThrLysGlyGlyGlnIleGly 1005
2461 GCTTCGGCAATGCTTCATTAATCAAGCGCAACGCGCGCAACAAACG 2510
1006 ValSerThrLysThrGlySerAspIleThrValAlaIleLysGlyGlnI 1022
2511 CAGTCGACGCTTCCGCG.....ACGCTAAGCAACGTAATTCATTT 2554
1022 yGlnThrThrAspAlaIleMetGlnThrLysAlaLysGlySerGlnP 1039
2555 CCGCACTCAACGCTAATGCTCCGCAAGCGCAATTAAGCGATTTTCATTT 2604
1039 heThrSerAsnGlyAspIleSerIleAsnValGlyGlnAlaHisIstY 1055
2605 GAAACGACGCGCTTACCGCAAA.....ATCAGCGCGCG 2639
1056 GluGlyAlaGlnPheAspAlaGlnLysGlyLysThrValIleAsnIaG 1072
2640 CAAGGATACGCGCATTAACCTTAAGACAGCAATGAGCGCTGCGCTCA 2689
1072 yGlyAspLeuThrLeu..... 1077
2690 GCACGGAATTAAGCAATTAACCTTGAACAGCGCAACATTAACATCAAT 2739
1077 1077
2740 TCGCCTATCGCCGACATGCGGAGGCGCAACGCGAGTCCGCAAGA 2789
1078 1082
2790 TCGCGCGCGCGCGCTTCCGCGCGCTTCCGCGCTTCCATTAATCCGTA 2839
1082 P..... 1082
2840 CACCGCACTTGGTAGAATCCGTTTCAACACGCTGACGGTAACGCG 2889
1083ThrHisSerGlnSerGlnSerAsn.....ValAsnGly 1093
2890 AAATTGAACGGTCAGGGGAACATTCCGCTTATGTCGGAACCTTGGCTA 2939
1094 SerAlaAsn..... 1096
2940 CCGCAGCGCAATTAAGCTGCGGAAAGTTCGAGAGCACTTACACT 2989
1097LeuLysValGlyThrThrProGlnSerLysAspIleG 1109
2990 TGGCGGTCAACATACCGCAACGCAACGCTGCAACGCTTCAATTAACG 3039
1109 yGlyGlyPheAsnAlaGlyThrHisHisSerLysGlnGlnThrThr 1125
3040GTAGTGAAGAAAGAACAAACCGCTGCGCAAAA 3077
1126 AlaLysValGlyThrIleThrGlySerGlnIleGlyLeuAsnAlaG 1142
3078 CTTAATTTCACTGCAAAACGAAACGATCGCGCGCGCTGCGCTT 3127
1142 yHisAsnLeuThrLeuGlnGlyThrHisLeuSerSer..... 1154
3128 ACCAATCATCCGCAAGAGCGGAGTTCGCTGCTAATCCGCTC 3174
1155GluGlnAspIleAlaLeuAsnAlaThrHisLysValAsp 1167
3175AAAGAAACAGAGCTTCCGA 3194
1168 LeuGlnSerAlaSerGlnHisThrGlnLysGlyAsnAsnLeuSerG 1184

3195 CAACCTCGCGCAAGCGCAAGAACCAAAACAGCGCGAAACAGACGCGC 3244
1184 yGlyValGlnAlaGlnPheGlyLysLysMetThr.....AspAspAla 1199
3245 AAAGCCTTACGCGCTGATTCGGCGCGCGCGGATGCGCTGAAACAGCA 3294
1199 eSerValAsnGlyLeu.....GlySerAlaGlnPheAlaIleGlyGln 1214
3295 GAAAGCTTGGCAACCGCGCGCGCGAGCAGC..... 3327
1215 AspGln.....LysSerValSerArgGlnGlyThrIleAsnAsnSerG 1230
3328GGGAAATGCTGCGCATTAATGAGCGGAGAGAG 3361
1230 yAsnLeuThrIleAsnGlyAsnSerValHisLeuGlnGlyAlaGlnVal 1247
3362 AGAAAAACGCGTGCAGCG..... 3381
1247 snSerLysAspThrGlnLeuThrSerGlnSerGlyAspIleGlnIleThr 1263
3381 3381
1264 SerAlaGlnSerThrAspIleLysAsnAsnIlePheGlyThrAspIleGlyP 1280
3382GATTAAGACACCGCTTGGCAAAACGCGCAAGCGGAAACCGCGC 3427
1280 eAsnGlyLysLysThrAsnAsnThrProLysGlyValIleThrGlnLysP 1297
3428 CGGCTACACCGCGCTTCCGCGCGCGCGCGCGCGCGCGGATTTGCCG 3477
1297 roAlaThrSerIleHisAsnIleGlyLysLeu.....LeuVal 1310
3478 CACTGCAACCCCAACCGCACCGCCCAACCGCAAGCGGACCTGATCAGCG 3527
1311 AsnValGlnAspGlnGlnLysThrSerHisGln..... 1321
3528 TTATGCCAATAGCGGTTGAGTGAATTTCCGCAACGCTCAACAGCTTT 3577
1322AsnAlaThrLeuGlnThrGlyThrLeuThrIleAsnSer 1334
3578 TCGCCCTACGACGACGATTAAGACCGCGTATTTCGGAAGACCGCGCGAC 3627
1335AsnLysAspLeuThrLeuSerGlyAlaAsnValIleAlaAsp 1348
3628 GCGGTTTGGCAACGCGCATCCGCGCAACCAACACTACGCTTGCAGA 3677
1349 SerVal.....ThrGlyAsnValGlyLysSerLeuAsnIleAlaSerGln 1363
3678 TTTCCGCGCTACCGCCCAACCAACGACCTGCGCAATCGGTATGCA 3727
1364LysGlnSerAsp.....ArgHisVal..... 1370
3728 AAAACCTGCGACGCGCGCTGCGCATCTGTTTGCACAAACGAGC 3777
1371ThrValGlyValAsnValGlyLysThrHisIstY 1381
3778 GAAACACCTTGCAGACGCGCATGCGCAACGCGGACGCGCTTCCACG 3827
1382 AsnAspProLysSerSerGlnValAsnLysThrAlaLys...AlaGly 1397
3828 GCGGCTTTTGGCAATAGCGCATGACAGCTTCTACATGCGCATCGAG 3877
1397 ySerLeuLeuGlnLysThrIleLysAspThrIleAspSerGlyIleLys 1414
3878 CGGCGCGCGGTTTATGACG.....GGCAGCTTTCACAGCGCATC 3918
1414 eSerThrAspAlaIleSerAspLysThrAsnSerLeuSerThrIle 1430
3919 GAGGCAAAATCCGCGCGCGCTGCTGCAATTACGCGATT..... 3957
1431 AlaAspLysThr.....GlyIleSerAspGlnIstY 1440
3958 .CAGGCACGATACGCGCGCTTCCG.....GGATTGCGCATGCAAC 4000

```

1440 rlysalatlyleaspginglypheylylysalatlylysalatlylysa 1457
1401 CGGACATCGGCGGACGCGGATTTTCACAAAGCGGATTCAC 4050
1457 snilevalthrgylalagluglyhsthrhlaasnlalspllelysa 1473
4051 GAAACGTCAT 4062
1474 ThrHisValasp 1477

seqname: SwissProt:40:OMPA_R1CNCN

seq_documentation_block:
ID OMPA_R1CNCN STANDARD: PRT: 2021 AA
AC 052657; P95591; P95592; P95593; P95594; Q52667; Q52668; Q52669;
AC 052670; Q52674;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Outer membrane protein A precursor (190 kDa antigen) (Cell surface
DE antigen) (rompa) (OMP A).
GN OMPA OR RCI1273.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Malish 7;
RX MEDLINE=94171067; PubMed=8125327;
RA Crocquet-Valdes P.A., Weiss K., Walker D.H., Barbe V.,
RT "Sequence analysis of the 190-kDa antigen-encoding gene of Rickettsia
RT conorii (Malish 7 strain).";
RL Gene 140:115-119(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RT Science 293:2093-2098(2001).
RN [3]
RP SEQUENCE OF 8-204 FROM N.A.
RC STRAIN-Indian tick typhus, M1, Malish 7, and Moroccan;
RX MEDLINE=97015921; PubMed=8662558;
RA Roux V., Fournier P.-E., Raoult D.;
RT "Differentiation of spotted fever group rickettsiae by sequencing and
RT analysis of restriction fragment length polymorphism of PCR-amplified
RT DNA of the gene encoding the protein rompa.";
RL J. Clin. Microbiol. 34:2058-2065(1996).
RN [4]
RP SEQUENCE OF 953-2012 FROM N.A.
RC STRAIN-Indian tick typhus, M1, Malish 7, and Moroccan;
RA Raoult D., Fournier P.-E., Roux V.;
RT "Phylogenetic analysis of spotted fever group rickettsiae by study
RT of the outer surface protein rompa.";
RL Submitted (DEC-1996) to the EMBL/Genbank/DBJ databases.
CC - FUNCTION: ELICITS PROTECTIVE IMMUNITY (BY SIMILARITY).
CC - SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC - PPM: GLYCOSYLATED (BY SIMILARITY).
CC - SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP B FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL: U01028; AAA17405.1; -
DR EMBL: AE008674; AAL03811.1; -
DR EMBL: U43794; AAB49549.1; -
DR EMBL: U43798; AAB49550.1; -
DR EMBL: U43806; AAB49551.1; -
DR EMBL: U45244; AAB49566.1; -
DR EMBL: U46918; AAB49663.1; -
DR EMBL: U83440; AAC35176.1; -
DR EMBL: U83443; AAC35179.1; -
DR EMBL: U83448; AAC35184.1; -
DR EMBL: U83453; AAC35189.1; -
DR InterPro: IPR003858; rompa_OMP.
DR Pfam: PF02708; rompa_OMP; 1.
KW Antigen; Repeat; Signal; Cell wall; s-layer; Glycoprotein;
KW Complete proteome.
FT CHAIN 1..38
FT SIGNAL 39..2021
FT DOMAIN 238..946
FT DOMAIN 1424..1528
FT VARIANT 60..60
FT VARIANT 76..76
FT VARIANT 86..137
FT VARIANT 126..133
FT VARIANT 953..954
FT VARIANT 1245..1245
FT VARIANT 1308..1308
FT VARIANT 1877..1877
FT VARIANT 10..10
FT VARIANT 92..92
FT VARIANT 126..126
FT VARIANT 137..137
FT VARIANT 157..157
FT VARIANT 368..368
FT VARIANT 374..374
FT VARIANT 640..640
FT VARIANT 669..669
FT VARIANT 793..793
FT VARIANT 803..804
FT VARIANT 809..823
FT VARIANT 898..908
FT VARIANT 908..985
FT VARIANT 985..1009
FT VARIANT 1009..1013
FT VARIANT 1013..1013
FT VARIANT 1182..1182
FT VARIANT 1314..1314
FT VARIANT 1451..1451
FT VARIANT 1624..1624
FT VARIANT 1628..1628
FT VARIANT 1872..1872
FT VARIANT 1875..1875
FT VARIANT 1875..1875
FT VARIANT 1879..1879
FT VARIANT 1936..1936
FT VARIANT 1965..1970
FT VARIANT 1997..1997
FT VARIANT 2021..2021
FT VARIANT 203328..327FC42D7CB24668 CRC64;
SQ SEQUENCE

Alignment_scores:
Quality: 231.00 Length: 1576
Ratio: 0.326 Gaps: 76
Percent Similarity: 44.924 Percent Identity: 19.670

```

Alignment_block:

us-09-303-518d-649 x OMPA_R1CNCN
Align seg 1/1 to: OMPA_R1CNCN from: 1 to: 2021

152 ACCATATCTATCGGACGATTTGCCGAAATAAAGCAGATTGGAGTCGGG 201

[illegible][illegible]

```
1828 .....CTGTGTTACAGCCCGCCGACGAGACCGCCTGCTGC 1867
      : : : : : : : : : : : : : : : : : : : : : :
1187 palaglnleuileuglnasnsersegllyasnspargthrlethrL 1204
      : : : : : : : : : : : : : : : : : : : : : :
1866 TTTCGGGGGAGAACATTTAAACGGCAAC..... 1896
      : : : : : : : : : : : : : : : : : : : : : :
1204 eu.....glyalasnileasprroaspasnaspraglylleval 1218
      : : : : : : : : : : : : : : : : : : : : : :
1897 .....ATCACGCAACAAACGGCAA 1916
      : : : : : : : : : : : : : : : : : : : : : :
1219 lleuasnservalthrAlaglylyslysleuthrlealeaglygly 1235
      : : : : : : : : : : : : : : : : : : : : : :
1917 ACTGTT.....TTCAGCGGCA 1935
      : : : : : : : : : : : : : : : : : : : : : :
1235 sthrphegllyAlahnslyseuglnasprllevalrheleysglyug 1252
      : : : : : : : : : : : : : : : : : : : : : :
1936 .....CCACACGGCAGCGCTACATTCATTAAAC 1965
      : : : : : : : : : : : : : : : : : : : : : :
1252 lysphegllythrAlaglythrPheasnThrThrasnillevalleu 1268
      : : : : : : : : : : : : : : : : : : : : : :
1966 GACCATTTGTCGCAAAAAG...GGCATTCTCGGGGAAATGCTGTG 2012
      : : : : : : : : : : : : : : : : : : : : : :
1269 AspillethrcllyleuileuglylalaThrThralaasnvalle 1285
      : : : : : : : : : : : : : : : : : : : : : :
2013 GGACACGACTGGATCAACCGCACATTTAACGGAA..... 2049
      : : : : : : : : : : : : : : : : : : : : : :
1285 upheleysAspAlaValaGlnleuthrclnThrghlyasnilleglyphei 1302
      : : : : : : : : : : : : : : : : : : : : : :
2050 ..AACTTCCAAATTAAGGGGACAGCGCGTTCGGCAATGT... 2094
      : : : : : : : : : : : : : : : : : : : : : :
1302 euasprheasnAlaLysasnlyThrValThrleuasnspasnvalasn 1318
      : : : : : : : : : : : : : : : : : : : : : :
2095 .....GCCAAGTGAAGGCGATGTCATTGACATCACGCCCAAC 2138
      : : : : : : : : : : : : : : : : : : : : : :
1319 valAlaglythrvallyllyasnthrlyglylThrasnsglythrleuL 1335
      : : : : : : : : : : : : : : : : : : : : : :
2139 AGTTTTGGTGTGCGACCGCATCAAGCCACAACTGTACAGCTGGG 2188
      : : : : : : : : : : : : : : : : : : : : : :
1335 eValleuglylaserasn.....LeuasnThrVala 1346
      : : : : : : : : : : : : : : : : : : : : : :
2189 ACTGAGCGGTCGTGACAAATGTGCGAAAAACATTACCGACGATAA 2238
      : : : : : : : : : : : : : : : : : : : : : :
1346 snnglylleAlaIleleuLysValaGlylAlaglyasnvalThrlealals 1362
      : : : : : : : : : : : : : : : : : : : : : :
2239 GTGATTGCTTCATTGACTAGACCGACATCAGCGCATCTGCTGC 2288
      : : : : : : : : : : : : : : : : : : : : : :
1363 .....glylyasnvallyllystleGI 1369
      : : : : : : : : : : : : : : : : : : : : : :
2289 CGAT.....CAGCTCATTTAA 2305
      : : : : : : : : : : : : : : : : : : : : : :
1369 yelulleleuglnlythrlylThrasnThrleuthrLeuProAlaHisshel 1386
      : : : : : : : : : : : : : : : : : : : : : :
2306 ATCTCAGAGGCTTGCACACTCAAGGC.....AAT 2337
      : : : : : : : : : : : : : : : : : : : : : :
1386 yleuthrthglyserlleasnlysthrlyglylnAlaLeuLysleasn 1402
      : : : : : : : : : : : : : : : : : : : : : :
2338 CTTAGTGAATGGCGATACAGCTTATAGTCAGCCACACAGCCACCA 2387
      : : : : : : : : : : : : : : : : : : : : : :
1403 pheMetasnlyglyserValserglyValaGlylThralaIaIaasnse 1419
      : : : : : : : : : : : : : : : : : : : : : :
2388 AAAGGCAACCTTAGCTGCTGGC.....A 2413
      : : : : : : : : : : : : : : : : : : : : : :
1419 rValaGlyaspIlethrThralaglylAlaThrserPheAlaserseVala 1436
      : : : : : : : : : : : : : : : : : : : : : :
2414 ATGCCCAAGCAACATTTAATCAAGCCACATTAACGGCAACATCGCT 2463
      : : : : : : : : : : : : : : : : : : : : : :
1436 smaIaLyslythr.....AlaThrleuglyllyThrThrserPhe 1449
      : : : : : : : : : : : : : : : : : : : : : :
2464 TCGGGAATGCTTCATTATCTAAGCGACGCGGTACAAAC...GG 2510
      : : : : : : : : : : : : : : : : : : : : : :
1450 Ala.....HisThrPheThrAsnThrGI 1457
```

```
2511 CAGTCTGACGCTTTCGGCAACGCTAAGCGAAACGTAAGCATTCGCAC 2560
      : : : : : : : : : : : : : : : : : : : : : :
1457 yAlaValThrleu.....AlaLysglyserIleThrserPheAla 1471
      : : : : : : : : : : : : : : : : : : : : : :
2561 TCACAGGTAATGCTCC.....CTACCGATAAGCAGTATTCATTT 2604
      : : : : : : : : : : : : : : : : : : : : : :
1471 yAsnValThrAlaThrserPheValAlaIaasnserAlaThrIleasnPhe 1487
      : : : : : : : : : : : : : : : : : : : : : :
2605 GAAGCAGC...CGCTTACCGGACAAATCAGCGGGCGCAAGATACGC 2651
      : : : : : : : : : : : : : : : : : : : : : :
1488 glyAsnserleuAlaPheasnserAsnIlethrly..... 1499
      : : : : : : : : : : : : : : : : : : : : : :
2652 ATTACACTTAAGACAGCATGAGCTGCTCCATCAGCAGCAATTAAG 2701
      : : : : : : : : : : : : : : : : : : : : : :
1500 .....SerGlyThrThrleu 1504
      : : : : : : : : : : : : : : : : : : : : : :
2702 GCAATTTAAACCTTGACACGCCACCATTTACATTCATTCGGCTATCGC 2751
      : : : : : : : : : : : : : : : : : : : : : :
1505 ..ThrleuglylAlaasnGlnvalThrlyThr..... 1514
      : : : : : : : : : : : : : : : : : : : : : :
2752 CACGATGGCGGCGGCGAAACCGCAGTCGACAGATGCGCGCGCG 2801
      : : : : : : : : : : : : : : : : : : : : : :
1515 .....GlyThrcllyserThrAsprThr..... 1522
      : : : : : : : : : : : : : : : : : : : : : :
2802 CGTTTCGCCGCTGCGCGCGTTCCTTATTCGCTTACACCGCAACTT 2851
      : : : : : : : : : : : : : : : : : : : : : :
1523 .....LeuthrleuasnThrThrPheaspIyA 1532
      : : : : : : : : : : : : : : : : : : : : : :
2852 CGGTAAATCCCGTTTCACACGCTGACGCTAACGCAAGCAAA...TTGAC 2898
      : : : : : : : : : : : : : : : : : : : : : :
1532 laAlaLysserglyllyasnIleleuIleLyssercllysetThrleuasp 1548
      : : : : : : : : : : : : : : : : : : : : : :
2899 GGTGAGGGAACATTCGCTTATGTCGACCTTCGCTGCTACCGCAGGA 2948
      : : : : : : : : : : : : : : : : : : : : : :
1549 leuSercllyAlaserasnleuAlaLeuValAlaThrAlaThrAspPhe 1565
      : : : : : : : : : : : : : : : : : : : : : :
2949 CAATTCAGCTGCGGAAAGTTCCGACGCACTTACACTTGCGCGTCA 2998
      : : : : : : : : : : : : : : : : : : : : : :
1565 pmetasnAsnIleSerProAsprThrlystlyThrvalIleaserAlaIuT 1582
      : : : : : : : : : : : : : : : : : : : : : :
2999 ACAATACCGGCAACGAACTGCAAGCTCGAACATTCGACGTAAGTGA 3048
      : : : : : : : : : : : : : : : : : : : : : :
1582 hrAlaglylyleuLysProThrProLysGlnasnvallystlethrIle 1598
      : : : : : : : : : : : : : : : : : : : : : :
3049 GGAAGAAGCAACAAACCGCTGCTC.....GAAACCTTAATTCAC 3089
      : : : : : : : : : : : : : : : : : : : : : :
1599 AsnasnspasnargPheValAspPheThrPheaspAlaserThrleuth 1615
      : : : : : : : : : : : : : : : : : : : : : :
3090 CCTGCAAAACGACACGTCGATGCCGCGCTTACCAACTCATCC 3139
      : : : : : : : : : : : : : : : : : : : : : :
1615 rleuPheAlagIuaspIleAlaIaIaglyVal..... 1625
      : : : : : : : : : : : : : : : : : : : : : :
3140 GCAAAGACGGGAGTCCGCTGCAATAATCCGTCAAAGAACAGAGCTT 3189
      : : : : : : : : : : : : : : : : : : : : : :
1626 ..lleaspGluaspPheAlaProglyllyProleu..... 1636
      : : : : : : : : : : : : : : : : : : : : : :
3190 TCCGCAACATCGGCAAGCGAAGCCAAAAACAGCGCGAAAAAGCAA 3239
      : : : : : : : : : : : : : : : : : : : : : :
1637 .....AlaasnIleProAs 1641
      : : : : : : : : : : : : : : : : : : : : : :
3240 CGCGCAAGCCTTGACGCGTATTCGGCGCGCGCGGATGCGTCAAA 3289
      : : : : : : : : : : : : : : : : : : : : : :
1641 nAlaIaIaasnIleLyslyserleuGluIleuMetGluAspAlaProAsng 1658
      : : : : : : : : : : : : : : : : : : : : : :
3290 AGACAGAAAGCGTTCCGACACCGCGCGGCGAGCGGGGAGAAATGTC 3339
      : : : : : : : : : : : : : : : : : : : : : :
1658 lyserAsp.....AlaArgGlnAla.....PheasnAsnPhe 1668
      : : : : : : : : : : : : : : : : : : : : : :
3340 GGCATTATTCAGCGAGAGAGAAAGAAACGGGTGACGCGGATTAAGA 3389
      : : : : : : : : : : : : : : : : : : : : : :
1669 GlyleuMetThrProleuglnlu.....AlaaspAlaThr 1680
      : : : : : : : : : : : : : : : : : : : : : :
3390 CACCGCTTGGGAAACAGCGGAAAGCGAAACCGCGGCTACACGCG 3439
```

```

1680  ||| |||
1680  rThHisLeu..... 1683
3440  CCTTCCCGCGCGCGCGCGCGCGCGCGGATTCGCAACATGCAACCC 3489
1684  .....MetGlnAsp 1686
3490  CAACCCGAGCCCAACCGCGCGCGCGCGCGCGCGGATTCGCAACCGCTTATGCCAATAG 3539
1687  ValValValProSer.....AspThrIleAlaValAlaAsnAsnGly 1700
3540  CGGTTGAGTGAATTTCCGCCACGCTCAACGCGGTTTCCGCGTACAG 3589
1700  nValValAla.....SerAsnIleSerSerAsnIleThrAlaLeuAsnA 1715
3590  ACGAATTAGCCCGCTATTGGC..... 3612
1715  laArgMetAspLysValIleAlaIleGlyProValSerSerGly 1731
3613  .....GAAGACCGCGCGCGCGCGCGCGGTTGGACAGCGCG...ATCCG 3650
1733  AspGluAspMetAspAlaLysPheGlyAlaTrpIleSerProPheValG 1748
3651  GGACACCAACACTACCTGCGCAAGTTCCCGCGCTACCGCAACAA 3700
1748  yAsnAlaThrGlnLysMetCysAsnSerIleSerGlyTyrLysSerAsp 1765
3701  CGGACCTCGCGCAATCGGTATGCAAGAAACCTCGCGCGCGCGCG 3750
1765  hTrpGlyGlyThrIleGlyPheAspGlyPheValSerAspLeuVal 1781
3751  ...GGCATCCGTTTTCGCGACACCGCGCGCGCGCGCGCGCTTGACGAC... 3795
1782  LeuGlyLeuAlaTrpThrArgAlaAspThrAspIleLysLeuLysAsnAs 1798
3796  ...GGCATCGCACTCGCGCGCGCGCGCGCGCGCGCGCGCGCTTTCG 3838
1798  nLysThrGlyAspLysAlaValIleGlnSerAsnIleLysSerLeuTyrG 1815
3839  GGCAATACGCGCATCGACAGGTTCTACATCGCGCATCGCGCGCGCG 3888
1815  LysLeuTyrSerValProTyrGlnLeuPheValIleAlaIleAlaSer 1831
3889  TTTAGAGCGCGCGCGCTTCACAGCGCATCGAGGCAAAATCCG... 3933
1832  TyrSerAsp.....AsnLysIleArgSerIly 1840
3934  ...CGCCGCGTGTGTCATTCAGGCGATTCAG...GCACGATACCGCGCG 3976
1840  sSerArgAlaValIleAlaThrThrLeuGlnThrValGlyTyrGlnThr 1857
3977  GTTTGGGCGGATTCGCGCATCGACCGCGCATCGCGCGCGCGCTATTTC 4026
1857  lAsnGlyLysTyrLysSerGlnSerIleTyrGlnLeu..... 1870
4027  GTCCAAAAGCGGATTCACCGCTAC.....GAAACGTCATATTCG 4067
1871  ....MetAlaGlyTyrThrTyrMetMetSerGlnAsnIleAsnLeu... 1884
4068  CACCCCGCGCTTCGATTCACCGCTACCGCGCGCGCGCGCGCGGATTC 4117
1885  ThrPro...LeuAlaGlyLeuArgTyr...SerThrIleLys...AspL 1898
4118  ATTCATTCACCAACCGCG...CAACACATTCGTCATTCACCGCT 4155
1898  ySerTyrLysGlnThrGlyThrThrTyrGlnAsnLeuThrValLysGly 1914
4156  TATTGAGCTGTCTATACCATGCGCGTTCGCGGCAAGTCGAACAG 4205
1915  LysAsnTyrAsnThrPheAspGlyLeuGlnAlaLysValSerSerAs 1931
4206  CGTCAATAC...GCCGTAATTCG 4225

```

```

1931  nIleAsnValAsnGlnIleValLeuThrProGlnLeuTyrAlaMetValA 1948
4226  CTCAGGATTTTCGCAAAACCGCGCGCGCGCGCGCGCGGATTCGCAACCGCGAA 4275
1948  sPTrAlaPheLysAsnLysValSerAla.....IleAspAlaArg 1961
4276  ATCAAGATTTTCACGCTGTCTTCACAGCGCGCGCGCGCGCGCGCGCA 4325
1962  LeuGlnIleMetThrAlaProLeuProThrAsnSerPheLys.....GI 1976
4326  ACTGGAAGCGCGCAACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4353
1976  nSerLysThrSerPheAspValGlyVal 1985

seq_name: SwissProt_40:OMP_RICRI
seq_documentation_block:
ID      OMP_RICRI      STANDARD;      PRT;      1654 AA.
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Outer membrane protein B precursor (168 kDa surface-layer protein)
DE      (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)
DE      (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE      antigen) (120 kDa outer membrane protein ompb); 32 kDa beta peptide].
GN      OMPB.
OS      Rickettsia rickettsii.
OC      Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC      Rickettsiaceae; Rickettsiinae; Rickettsia.
OX      NCBI_TaxId=783;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-R;
RX      MEDLINE=92167802; PubMed=1724278;
RA      Gilmore R.D. Jr., Joste N., McDonald G.A.;
RT      "Cloning, expression and sequence analysis of the gene encoding the
RT      120 kD surface-exposed protein of Rickettsia rickettsii.",
RT      rickettsii is encoded by an unusually long open reading frame:
RT      evidence for protein processing from a large precursor.",
RT      Mol. Microbiol. 5:2361-2370(1991).
RN      [2]
RP      SEQUENCE OF 279-1654 FROM N.A.
RC      STRAIN-R;
RX      MEDLINE=90136087; PubMed=2515418;
RA      Gilmore R.D. Jr., Joste N., McDonald G.A.;
RT      "Cloning, expression and sequence analysis of the gene encoding the
RT      120 kD surface-exposed protein of Rickettsia rickettsii.",
RT      Mol. Microbiol. 3:1579-1586(1989).
CC      -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC      STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC      VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC      -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC      LAYER WITH HEXAGONAL SYMMETRY.
CC      -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: X16353; CAA34403.1; -
DR      InterPro: IPR003858; rOmpA_rOmpB.
DR      Pfam: PF02708; rOmpA_rOmpB; 1.
KW      Antigen; S-layer; Cell wall.
FT      CHAIN          1..1333
FT      DOMAIN        1334..1654
FT      DOMAIN        1181..1188
FT      DOMAIN        1654 AA; 168184 MW; D7AB70PB7087618 CnC64;
SO      SEQUENCE

```

alignment_scores: Length: 1603

Quality: 224.50 Gaps: 68

Ratio: 0.316

Percent Similarity: 44.292 Percent Identity: 19.152

alignment_block:

US-09-303-518D-649 x OMPB_RICRI ..

Align seg 1/1 to: OMPB_RICRI from: 1 to: 1654

```

244 GCGAATTCATGCAAGCCGATGATGTTTTCGTGTCGCG 293
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
130 Gilythrleuthr.....lletthnglynglyvalthr 142
294 TAACGGCGTGGCGCATTTGGTGGCGATCATATATGACG 343
    : : : : : : : : : : : : : : : : : : : :
142 nlaaglinalaialthrlysnalaglnasnaivalalgn 159
344 ATACGGC.....GCTATTAACAAGCTT 366
    : : : : : : : : : : : : : : : : : : : :
159 snasnglyalaialaspasnaspleuysglyvalglyar 175
367 GATTTCGTCGGAAGAGAATCCGATCAACATCCTTTACT 416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
176 Asphneglyalaproalaserthr.....leuvalpheas 187
417 AATTGTGAACGAAATATTAAGA..... 444
    : : : : : : : : : : : : : : : : : : : :
187 nleuasnprothrtnglnlysalaproleuileu glyasp 204
445 .....GGGACTAAAGCCATCCTTATGCGCGC 471
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
204 lavalilleasnglyvalasnglythrleuasnaival 219
472 GATTTCATATGCCCGTTGATTAATTTGTACAGATCGA 521
    : : : : : : : : : : : : : : : : : : : :
220 ...Pheileglnvalserasnlyserphealathr..... 231
522 TGAATGACCACTTATGATGGCGGAATATATC.....GATC 562
    : : : : : : : : : : : : : : : : : : : :
231 llysalalleasnllealaspglynglylletlepheas 248
563 AAAATATTAACCTGACCTGTTCATATGGCGCAGCA 612
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
248 laasnasn..... 250
613 CGATCTGATGAAGATGACCCCAATACCGGAAGTTCA 662
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
251 .....A 251
663 AAGTGGTATTCCTTGGCTGTCGCAATACCTTGCACAA 712
    : : : : : : : : : : : : : : : : : : : :
251 aasrthleuasnlleuaglnalaglythrthrleasn 268
713 CAGTGGT...GGCAGATCACTTAGTGTGAAAAAATTA 759
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
268 hraspelythrnglyargleuvalleuuser..... 281
760 CCATATGGTTTTTACCAACAGAGGTGTCATTTGGCAG 809
    : : : : : : : : : : : : : : : : : : : :
282 Alalaathrasnphenasnlierthglyserleuagly 298
810 AATGTTTATTCATGATGCCAAAAGCAAAAGTGTATTA 859
    : : : : : : : : : : : : : : : : : : : :
298 yaalileglurpheasntrvalalavalasp glyln 315
860 TCGAAGCGGCAACCCCTATATAGAAAAGCAATGGCT 909
    : : : : : : : : : : : : : : : : : : : :
315 laglyalalalalalavalilleglythrhrasnng 331
910 CGTAAAGATGTTGTTCTATGATGAATCTTTCGAGAT 959
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
332 .....Aaglyphevalvalserva 338

```

```

960 ATTTCAGACACACCTCAAAATGGGAAATCTTTTAAC 1008
    : : : : : : : : : : : : : : : : : : : :
338 l.....aspasnglyvalalalathrilleaspol 350
1009 .....AATGCACAGCAAAATC 1026
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
350 altyralalayspmetvalileginseralalasnali 366
1027 AATGCCAAACATGAACACAAATTCCTGCTAATAGAT 1076
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
367 Asnphearghis..... 370
1077 CGTTCAATGTTTAAATGTTCTTTATCCGACAGACAG 1126
    : : : : : : : : : : : : : : : : : : : :
371 .....ilevalaspval..... 374
1127 ATCATGCTGCAGGTGCTGCACACGTTATCGACC..... 1161
    : : : : : : : : : : : : : : : : : : : :
375 .....glyalasp glythrthrthralapheus 389
1162 AGACGTAATGAGAGAAATATTCCTTTATTTAGCAAG 1206
    : : : : : : : : : : : : : : : : : : : :
390 Thrllthrclinaspsersnphneglyasnthrph 406
1207 .....GGCAATTGATCTTACCAGCAATC..... 1233
    : : : : : : : : : : : : : : : : : : : :
406 aalaglnllelyvalproasnalalierthleuthr 423
1234 .....ATCAAGT.....GCTGAGAGATTATTTCCA 1266
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
423 lypasplaserasnproglyasnthrhalaglyall 439
1267 GATTTTACGGTTCGCTGAAATTAACGAAACTTGGCA 1316
    : : : : : : : : : : : : : : : : : : : :
440 Asnglythrleugluseralaseralasp.....Alas 452
1317 TCATTCAGTGAGACACGACGTTACTTGGAAATACG 1366
    : : : : : : : : : : : : : : : : : : : :
452 lalavalthrnsnlnlierthralalle..glualas 468
1367 ACGACCGCTGTCAAAATCGCAAAAGCAGCTGCAAG 1416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
468 alvalcrlneuser.....glythrhisalalaleu 480
1417 GGGAAAACCAAGGCTCATC...AGCGTGGCGAGCT 1463
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
481 leu glyasnalaglyserllepheylsleu alasp 496
1464 GATTCAGCAGCAGACGATTAAGCAAAACACCTTTAG 1513
    : : : : : : : : : : : : : : : : : : : :
497 .....Asnglylyvalasnainglnthr alaleu 507
1514 GCTTGGTCAGCGCAGGAGCTGACGATGATGCGAT 1563
    : : : : : : : : : : : : : : : : : : : :
507 llyglyalaleu alalaglythrlierthleuaspol 523
1564 AACCCGAC..... 1572
    : : : : : : : : : : : : : : : : : : : :
524 Thrlyasprileglyasnalaglyalalalaleucl 540
1573 .....AAACTGTATT 1582
540 flenalalasnasp alalalyslythrleuthrleu 557
1583 TCGGCTTTCGCGCGGAGCGTTGGATTA.....AAC 1626
    : : : : : : : : : : : : : : : : : : : :
557 leglyalaglyglylthrleaspleu lnalasnngly 573
1627 TCGTTCACCGTATTCAAAT..... 1647
    : : : : : : : : : : : : : : : : : : : :
574 lysleuthrserthrnglnasnasnillevalalasp 590

```

1648ACGATGAAGGCGCATGATGTCACCAACATCAAGACAAAGAAAT 1693
590 ealAtHrAspGlnThrGlyValAlaSerSerLeuThrAsnAlaG 607
1694 CCACCGTACCATTAACGCAATTAAGATATGCTACCAACCGCAATTAAC 1743
607 InHrLeuThrIleAsnGlyLys.....IleGlyThrIleGlyAlaAsn 621
1744 AAC.....ACCTGATAGCAAAAAAGAAATGCTTA 1775
622 AsnLysThrLeuGlyGlnPheAsnIleGlySerLysThrValLeuSe 638
1776 CAACGGTGGTTGGCGAGAAAGAT..... 1800
638 rAsnGlyAsnValAlaIleAsnGlnLeuValIleGlyAsnAspGlyAlaVal 655
1801ACGACCAAAACGAC 1815
655 aGlnPheAlaHsAspHrTyLeuIleHrArgThrAsnAlaAla 671
1816 GGGCGGCTCAACCTGTTTACGAGCCCGCGAGAGACCGACCTGCT 1865
672 GlyGlnGlyLysIleIlePheAsnProValValAsnAsnGlyThr..Th 687
1866 GCTTCGCGCGGACCAATTAAGCGCAATCAGCAAAACGCA 1915
687 rLeuAlaIleGlyThrAsnLeuGly.....SerAlaThrAsn... 699
1916 AACGTTTTCAGCGGACCAACACCGCAACGCTTACATCATTTAAC 1965
699 699
1966 GACCATGTGTCAGAAAAGAGGATTCCTCGCGGAAATCGTGCGA 2015
700ProLeuAlaGlnIle..... 704
2016 CAACGACTGATCAACCGCATTTAAAGCGAAACTTCCAAATTAAG 2065
705AsnPheGlySerLys 710
2066 GCGGACAGCGGCTGTTTCCGCAATGTCGCAAAAGTGAAGCGGATG 2115
710 LyValAsnValAspHrValLeuAsnValGlyLys..... 721
2116 CATTTGACATCAACGCGCAAGAGTTTGGTGCACCGCATCAAG 2165
722GlyValAsnLeuTyAlaTh 728
2166 CCAACACATCTGACAGTTCGACTG.....ACGG 2197
728 rAsnIleHrThrThrAspAlaAsnValGlySerPheValPheAsnAlaG 745
2198 GTCGCAAAATGTCGAAAAAACCAT..... 2226
745 LyGlyThrAsnIleValSerGlyThrValGlyGlnGlnLysLys 761
2227ACGACGATTAAGTGAATGCTTCAATGCTAA 2258
762 PheAsnThrValAlaLeuGlnLysGlnLysThrValLysPheLeuGlyAla 778
2259 GACCGACATCAACGCAATGTCATGCGATGACGATCAACGCTTAAATC 2308
778 nAlaThrPheAsnGlyAsnThrIleAlaAlaAsnSerThrLeuGlnI 795
2309 TCACAGGGCTTGCACACTCAACGCAATCTTACTGCAATAGCGATACA 2358
795 IleGlyLysAsnTyThrAlaAspCysValAlaSerAlaAspGly..ThrG 811
2359 GCTTATACAGTCAGCCACACAGCCCAAAAGCAACCTTACCCTCGT 2408
811 yLleValGlnPheValAsnThrGlyProIleHrValThrLeuAsnLysG 828
2409 GGGCAATGCCAAGACATTTAAT.CAAGCCATTAAAGCGCACACA 2457

828 InAlaIleProValAsnAlaLeuLysGlnIleHrValSerGlyProGly 844
2458 TCGGCTTCGGGCAATGCTTCAATTAATCTAAGCAACAC..... 2496
845 AsnValAlaIleAsnGlnIleGlyAsnAlaGlyAsnHisHisGlyAlaVala 861
2497GCCGTCAAAAGCGAGCTG..... 2517
861 lHrAspHrIleAlaPheGlnAsnSerSerLeuGlyAlaValAlaPheL 878
2518ACGCTTCCGGCAC 2532
878 euProArgGlyIleProPheAsnAspAlaGlyAsnThrMetProLeuThr 894
2533 GCTAAGCAACGTAACCATTCGCGACTCAACGCTTAATGCTCCCTAGC 2582
895 lIleLysSerThrValGlyAsnLysThrAlaLysGly...PheAspValPr 910
2583 CGATAAGCGATATTCATTTGAAGAGCGCGCTTACCGGACAAATCA 2632
910 oSerValValAlaLeuGlyValAspSerValIleAlaAspGlyGlnValI 927
2633 GCGGC.....GCCAAGATACGCA 2652
927 lGlyAspGlnAsnAsnIleValGlyLeuGlyLysSerAspAsnGly 943
2653 TTACACTTAAGACAGCGAATGAGCGCTGCCGTCAGGACGCAATAGG 2702
944 lIleIleValAsnAlaThr.....ThrLeuTyAlaGly.....IleSe 956
2703 CAATTTAACCTTACAAACGCGCATTAATCACTCAATTCGCTATCGCC 2752
956 rThrLeuAsnAsnAsnGlnGlyThrValThrLeuSerGlyVal...P 972
2753 ACGATGCGCGCGGCGCAAAACGCGCGCGACAGATGCGCGCGCC 2802
972 rAsnThrProGlyThrValTyreGlyLeuGlyThrIleGlyAlaSer 988
2803 CGTTCGCGCGCTGCGCGCTTCCCTAATTCGCTTACACCGCAACTTC 2852
989 LysPheLysGln.....ValThrPheHrThrAs 998
2853 GGTGAATCCCGTTTCAACGCTGACGTAAC..... 2886
998 pTyraAsnAsnLeuGlyAsnIleIleAlaThrAsnAlaThrIleAsnAsp 1015
2886 2886
1015 lYValThrValThrThrGlyGlyIleAlaGlyIleGlyPheAspLys 1031
2887GGCAATTAAGGTCAGGACGTAACATTCGCTTATGTCGA 2927
1032 lIeThrLeuGlySerValAsnGlnLysAsnValAlaPheAlaAspG 1048
2928 ACTCTTGGCTACCGGACGACAAATTAAGTGCAGGAAGTCCGAG 2977
1048 yLleLeuSerAsnSerThrSerMetIleGlyThrThrLysAlaAsnSng 1065
2978 GCACCTACACCTTGGC.....GTCAACATATCCGGAACGAA 3015
1065 lYThrValThrTyreGlnLysAlaPheValGlyAsnIleGlyAspSer 1081
3016CCTGAAGGCTCGAACAATTCAGCTAGTGAAGGAAGAACAA 3059
1082 AspThrProValAlaSerValArgPheThr.....GlySerAspSe 1095
3060 CAACCGCTGTCGGAACCTTAATTCACCGCTGCAAAAGCAACAGCTG 3109
1095 r.....GlyAlaGlyLeuGlnGlyAsnIleTySerGlnValIle 1109
3110 ATGCGGCGCTGCGTTACCACTCAACGCAAAAGCGCGAGTTCGCC 3159


```

3205 laGlyGluAspTyrGlnAlaGlnAlaThrIleAspValGlyGln 3221
3247 AGCTTGAC...GGCTGATGCGCGGG... 3273
3222 ThrLysAspProAlaArgLeuGlnValGlyGlyValLysGlyThrLe 3238
3274 ...CGGATGCGCTCGAAAAGACAGAAAGCGTTGGCAACGGCGCGGC 3319
3238 uAsnGlnAspAlaAlaGlnAlaThr...ValValGlnArgAnLysHisT 3254
3320 AGCGAGCGCGG...GAAATGTCGCAATTATGAGCGGAGAGAG 3363
3354 rPalaglyGlyGlySerGln.PheSerValAlaGlyLysSerLeuLys 3270
3364 AAAAAGSGTGCAGCGGATTAAGACACCGCTTGCGAAACAGCGGCA 3413
3270 sLysAsnGlnValArgProValGlnThrProThr...ProAspValAla 3286
3414 AGCGAAACCGCGCGCTACACCGCTTCCCGCGCGCGCGCGCC 3463
3286 spLysProProSerArg...ProThrThrProProAlaSerPro 3299
3464 GCGGGATTTCGCGCACTGCAACCCAGCGGAGCGGCAACCGCAGCGC 3513
3300 GlnProIleArgAlaThrValGlnValSerSerProProValSerVa 3316
3514 GACCTGATCAGCGCTTATGCCA...ATAGCGGTTTGAGTGAATP 3354
3316 lAlaThrValGlnValValAlaProArgProLysValGlnThrAlaGln 3333
3555 TTCGCCACGCTCAACACGCGTTTCGCGCTAGACGAGCATTAAGACGG 3604
3333 euProProArgProValAla...AlaGln 3341
3605 TATTTGCCGAGACCGCGCGAGCGCTTGACAAAGCGGATCCGCGAC 3654
3342 ValValProValThr...Pr 3347
3655 ACCAAACCTACCTTCGCAAGAT... 3679
3347 oProLysValGlnValAlaLysValGlnValAlaProArgProLysValG 3364
3680 ...TCGCGCGCTACCGCGCAACAAACGACCTTCGCGCAATCGTATGC 3724
3364 luthrAlaGlnProLeuProProArgProValAlaAlaGlnLysVal... 3379
3725 AGAAAACTCGGACGCGCGGCGTGGCATCTTTTCGCAACAGCGG 3774
3380 ...ThrThrProAlaValGlnProGlnLeuAlaLysValGlnThrValG 3395
3775 ACCGAAACACCTTCGACAGC...G 3797
3395 nProValLysProGlnThrThrLysProLeuProLysProLeuProValA 3412
3798 CATCGCAACTCGG...CAGCGTTGCCACG 3826
3412 lAlysValThrLysAlaProProProValValGlnThrAlaGlnProLeu 3428
3827 GCGCGCTTTTCGCGCAATACGCAATCGATTCATCGCATCGCATCAGC 3876
3429 ProProValLysProGlnLysAlaThrProGlnProValAlaGlnValG 3445
3877 GCGGCGCGGCTTTTACGACGCGGAGCTTCAGACGCGCATCGAGAGCA 3926
3445 yLysAlaThrVal...ThrThrVal.GlnValGln 3455
3927 AATCGCGCGCGGCTGCTCATTCAGGCAATTCAGGACATACCGCGCGC 3976
3456 SerAlaProProLysProAla... 3462
3977 GTTTCGCGGATTCGCGATCAACGACGATCGCGGCGCATATTTTC 4026
3462 ... 3462

```

```

4027 GTCCAAAAAGCGATTACCGCTACGAAACGTCATATTCGCCACCCCGG 4076
3463 ..... 3473
4077 CCTTCATTCAACCGCTACCGCGCGGCGCATTAAGCAGATTATTCATCA 4126
3474 ProLysProLysProLysProLysAlaGlnArgPro.....L 3486
4127 AACCGCGCAACATTCATTCATCGGCTATTGAGCGCTGCTTACC 4176
3486 ySProGlyLys.....Thr 3490
4177 GATGCGCTTCGCGCAATCCGACACGCGTCAATACCGGATATTCG... 4224
3491 ThrProLeuSerGlyArgHisValValGlnGlnGlnValGlnValLeuG 3507
4225 .....GCTCAGATTTCGCGAAACCGCGAGTGGGATGCGC... 4263
3507 nArgGlnAlaSerAspLeuAsnThrLysSerLeuProGlyGlyLysL 3524
4264 .....GTAACGCGCAATCAAA 4281
3524 euProLysProValThrValLysLeuThrAspGlnValAsnGlyLysProGln 3540
4282 GGTTCACGCTGCTCCTCCACGCTGCGCGCGCAAA.....GCCCGCA 4325
3541 ThrTyrThrIleAsnArgArgGlnAspLeuMetLysLeuAsnGlyLysVa 3557
4326 ACTGGAAGCGGCACACACGCGCGCATCAATTAAGCTACCGC 4368
3557 lLeuSerThrLysThrThrLysGlyLeuGlnGlnThrPheArg 3571

seq_name: SwissProt_40:FIG2_YEAST

seq_documentation_block:
ID FIG2_YEAST STANDARD; PRT; 1609 AA.
AC P25653;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Factor induced gene 2.
GN FIG2 OR YCR089W OR YCR89W OR YCR1102.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NBLTaxID=4932;
RN [1]
RP
RX MEDLINE=92397594; PubMed=1523889;
RA Wilson C., Grisanti P., Frontali L.;
RT "The complete sequence of a 6146 bp fragment of Saccharomyces
RT cerevisiae chromosome III contains two new open reading frames."
RL Yeast 8:569-575(1992).
CC
CC FUNCTION: REQUIRED FOR EFFICIENT MATING.
CC -!- INDUCTION: BY MATING PHEROMONES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X59720; CAA4254.1;
DR PIR: S19504; S19504.
DR PIR: S25345; S25345.
DR SGD: S000065; FIG2.
SQ SEQUENCE 1609 AA; 166049 MW; 7D66AD7F85A7B852 CRC64;

```

alignment_scores:

Quality: 215.00 Length: 1187
 Ratio: 0.378 Gaps: 55
 Percent Similarity: 47.936 Percent Identity: 20.388

alignment_block:
 US-09-303-518D-649 x FIG2_YEAST ..

Align seg 1/1 to: FIG2_YEAST from: 1 to: 1609

```

1553 ATATATAGTTCAACCCGACACACTCTATTTCGCTTCGGCGGCGACGT 1602
      ||| :|||:| | | | | :|||:| | | | | :|||:| | | | |
159 lIseSerGIserThrSerProIySerleuGIserPheAspTr..... 213
1603 TTGGATTAAACGGCGCATTCGCTTCCTCCACCGTATTCAAAATACCGA 1652
      ||||| :|||:| | | | | :|||:| | | | |
214 .....Thrgly...ThrIleThrSerIySerProSerProS 226
1653 TGAAGGGCGCATGATTGTCACACACATCAAGACAAAGAAATCCACGTTA 1702
      :|||:| | | | | :|||:| | | | | :|||:| | | | |
226 erSerIyAsnSerAsnGIserThrSerleuSerProIySerleu 242
1703 CCATTACAGGCATTAAGATATTGCTACACCGCGCAATA.....ACAAC 1746
      :|||:| | | | | :|||:| | | | | :|||:| | | | |
243 SerSerSerSerGIyAspIleuIleleuSerSerThrIleGlnAlaThr 259
1747 ACCTTGATAGCAAAAGAAATGCTCCTACACGCTTGCTTGCGCGAGA 1796
      | :|||:| | | | | :|||:| | | | | :|||:| | | | |
259 rAsnAspGIserIySerThrIleProThrIleuValAspAlaThrSers 276
1797 AGATACGACCA.....AACGACAGGGCGGCTCA 1825
      :|||:| | | | | :|||:| | | | | :|||:| | | | |
276 erLeuProThrIleuArgSerSerSerMetalapProThrSerGIySer 292
1826 ACCTTG.....TTTACGACCCGCGCGACGACGACGCA..... 1858
293 AspSerIleSerHisAsnPhetSerProSerIySerThrSerGIyAs 309
1859 .....CCTGCTGCTTCGCCGCG 1877
309 nTyAspValIleuThrSerAsnSerIleAspProSerleuPhe..... 323
1878 AACAAATTAAACGGCACATCAGCAAAACAAACGCAAACTGTTTTC 1927
      |||||:| | | | | :|||:| | | | | :|||:| | | | |
324 .....ThrThrThrSerGIyIySerSerThrGlnIleuSerSer 336
1928 GCGGACACACACACGCGACGCTTACATTTAAAGCACCATTTGTCG 1977
      :|||:| | | | | :|||:| | | | | :|||:| | | | |
337 leuAsnArgAlaSerIySerGIuThrValAsnPhetTr..... 349
1978 CAAAAAGAGGCAATTCCTCGCGGGGAAATCGGTGGGACA..... 2017
      ||| :|||:| | | | | :|||:| | | | | :|||:| | | | |
350 .....AlaSerIleAlaSerThrProPheGIyThrAspSerAlaT 363
2018 .....ACGACTGGATCAACCGGCA..... 2037
363 hrSerIleuIleAspProIleSerSerValGIySerThrAlaSerSerPhe 379
2038 .....TTTAAGCGGAAACTTCACAAATTAAGCGGACAGCGGTGT 2081
      :|||:| | | | | :|||:| | | | | :|||:| | | | |
380 ValGIyIleSerThrAlaAsnPhetSerThrGlnGIyAsnSerAsnIyVa 396
2082 TTCGCGCAATGTTGCCAAAGTAAAGC.....GATGGCAT..... 2118
      | :|||:| | | | | :|||:| | | | | :|||:| | | | |
396 lProIuSerThrAlaSerGIySerSerGIyIyGlnAspTrSerSers 413
2119 .....TTGACAAACAGCGCCACAGCATTTTGGTTCGCGACG 2157
      |||||:| | | | | :|||:| | | | | :|||:| | | | |
413 erSerleuProIeSerGIuThrThrIyValIleAsnTrIleAsn 429
2158 CATCAAGCCACACAAATCTGTACACGTTGCGAC.....TGGACGGTCT 2201
      |||||:| | | | | :|||:| | | | | :|||:| | | | |
430 ThrGlnGIySerValIleThrSerThrIleSerProAlaTyValSerThral 446

2202 GACAAATGTGTGAAAAACATTACGACGATAAAGTATGCTTCAT 2251
      |||||:| | | | | :|||:| | | | | :|||:| | | | |
446 arThyIyThrValAspGIyValIleThrGIuThrIyValThrIyPysPro 463
2252 TGACTAAGACC.....GACATCAGCGGCAATGTGATCT 2286
      |||||:| | | | | :|||:| | | | | :|||:| | | | |
463 euThrGIuThrIySerSerGIuAlaIleGIyValSerSerSerIleSer 479
2287 GCCGATCAGCCTCATTTAAATCTCACAGGCTGCGACATCAACGCA 2336
      :|||:| | | | | :|||:| | | | | :|||:| | | | |
480 ValProIuAlaSer...SerPheSerGIySerSerIleleuSerSers 495
2337 TCTTAGT.....GCAATG 2350
      | :|||:| | | | | :|||:| | | | | :|||:| | | | |
495 nSerSerThrIleuAlaIleAsnAsnValProIuSerThrAlaSerG 512
2351 GCGATACACGTTAT..... 2364
      |||||:| | | | | :|||:| | | | | :|||:| | | | |
512 lySerSerGIuThrGIuAspTrSerSerSerSerleuProIeSerGIu 528
2365 .....ACAGTCAGCCACACGCGCCCAAAAGCGCACCTTG 2402
      :|||:| | | | | :|||:| | | | | :|||:| | | | |
529 ThrThrIyValValIleAsnThrIleAsnThrIleGlnIySerValThrS 545
2403 CCTCGTGGCGCAATGCCACAGACATTTAATCAAGCACATTAACGGC 2451
      | :|||:| | | | | :|||:| | | | | :|||:| | | | |
545 rThrThrSerProAlaTyValSerThrAlaThrIyThrValAspGIy 562
2452 .....AACACATCG 2460
562 alIleThrGIuThrValThrIyPysProIeThrGIuThrIySerGIu 578
2461 GCTTCGCGCAATGCTTCATTTAATCTACGACGCGGTACAAAC... 2508
      |||||:| | | | | :|||:| | | | | :|||:| | | | |
579 AlaIleGIyIleSerSerSerThrIleSer.....AlaThrGIuThrS 593
2509 .....GGCAGTCTGACGCTTCGCGCACGCTTAAGCAACGTAAGCC 2551
      :|||:| | | | | :|||:| | | | | :|||:| | | | |
593 rIySProSerSerIleleuThrGIuGIyIleSerThrIleuIleuSerA 610
2552 ATTCGCGACTCAACGT.....AATGTCCTCCATGCGCATTAAG 2589
      |||||:| | | | | :|||:| | | | | :|||:| | | | |
610 spAlaThrIlePheIySGIyThrGIuThrIleAsnThrHisIleuThrGIu 626
2590 GCAGTATTCATTTGAAACGACGCGCTTACCAGCAAAATCAGCGGGG 2639
      :|||:| | | | | :|||:| | | | | :|||:| | | | |
627 SerThrSerIleThrGIuProIleThrIyPheSer.....GI 638
2640 CAAGGATACGCGATTACACTTAAGACAGCGAATGAGCGTGGCG... 2685
      :|||:| | | | | :|||:| | | | | :|||:| | | | |
638 yThrSerAspSerPheThrIeIySerThrSerGIuValAspIleAlaSerS 655
2686 .....TCAGGACGGAATTAAGCAATTTAAAC 2712
      |||||:| | | | | :|||:| | | | | :|||:| | | | |
655 erLeuSerSerIyProAsnPhetSerSerSerGIuGIySerThralaTr 671
2713 CTTGACAAAGCCACCATTAACATTCGCGCATTCGCGACGATCGGC 2762
      :|||:| | | | | :|||:| | | | | :|||:| | | | |
672 lIleThrAsnSerThrValThrPheIySerThrSerIySerProSer... 687
2763 AGGGGGCGCAACCGCGCAGTGCAGATGCGCGCGCGCGCTTCGCGC 2812
      :|||:| | | | | :|||:| | | | | :|||:| | | | |
688 .....ThrSerValSerAsnProThrGIuAlaSerGIu 699
2813 GTTCGCGCGCTTCCTATTATTCGTTACACGCGCAACTGCGTGAATCG 2862
      :|||:| | | | | :|||:| | | | | :|||:| | | | |
699 lValSerSerSerValAsnSerleuThrAspPheThrSerAsnSerThr 715
2863 CGT..... 2865
716 GIuThrIleAlaValIleSerAsnIleHisIyThrSerSerAsnIyAs 732
2865 ..... 2865

```

732 pTySerLeuThrThrThrGlnLeuThrSerGlyLysGlnThrLeuVal 749
2866 ..TTCAACAGCGCTGACGGTAACGGCAATTAACGGCTGACGAGCAATTC 2913
749 aLeuSerThrValThrThrThrValAlaSerGlnAlaThrGlnThr 765
2914 CGCTTTATGTCGAACTTCGGCTACCGGAGCAAAATTAACCTGAC 2963
765 rThrTrpCysProAlaSerSerThrLeuAlaThrThrSerThrSer 782
2964 GG..... 2965
782 yStrLeuValLeuThrThrGlnValCysSerHisSerGlnCysThrPro 798
2966 AAAGTTCGAAAGCACTACCTGACCTGGCGTCAACATACCGGCAACAA 3015
799 ThrValThrSerValThrAlaThrSerThrThrThrLeuLeuLeu 815
3016 CCGCAAGCGCTCGAA.....CAATGACGATGAGGAGAAAGACA 3058
815 rThrSerSerThrThrValLeuSerSerThrValSerGlnValAla 832
3059 ACAAAACCGCTCCGAAACCTTATTCACCTCGCAAAACGAAACACGTC 3108
832 snProAlaAlaSerGlu.....ValThrIleAsnThrIleVal 845
3109 GATGCGCGGCGTGGCTTACCAACTACCGCAAGACGCGGAGTTCCG 3158
845 rAlaThrSerGlnAlaThrSerThrThrGlnValSerAlaThrSer 862
3159 CCTGCATATCCGGTCAAGAAC..... 3181
862 lAlaThrAlaThrAlaSerGlnSerSerThrThrSerGlnValSer 3182
3182AAGAGCTTCCGACAACTCGC 3204
879 SerGlnThrThrSerThrThrGlnThrGlnAsnThrThrThrThr 3204
3205 AAGCGACAGCCA.....AAAACAGCGGAAAGAACACACG..... 3241
895 rLeuLeuPheProAlaLeuSerThrGlnThrLeuThrValVal 3242
3242 ..CGCAAGCGCTG.....ACGCGCTGATTCGGCGCGGCGCATG 3280
912 eTrnGlyThrLeuThrLeuThrThrGlnValCysSerHisSerLys 3281
3281 CCCTCGAAAG.....ACAGAAAGCGTTCGGAACCGCGCGGACGCA 3324
928 sValProThrValThrThrGlnValThrSerLysGlyThrPro 3325
3325 GCGCGGGAATGTCGGCATTCGACGCGGAGAGAGAAAGAAAGCGGT 3374
945 AaGlnHisSerSerGlnThrLeuGlnThrGlnAlaValGlnValThr 3424
3375 GCAAGCGGATTAAGACACCGCTTCGGAACAGCGGAGCGGAAAC 3423
961 uSerSerHisGlnThrValThrMetSerThrGlnValCysSerHis 3444
3424CGCGCGGCTACACCGCGCTTC 3444
978 lCysThrProThrValThrSerValGlnMetArgSerThrProPhe 3445
3445 CCC..... 3447
995 ProTrpLeuThrSerThrSerSerSerSerLeuAlaSerThrLys 1011
3448 ..CGGCGCGCGCGCGCGGATTCGCAACTGCAACCCCAACCGC 3496
1011 sSerSerLeuGlnAlaSerSerGlnMetSerThrPheSerValSer 1028
3497 AGCCCAACCG..... 3507
1028 lSerLeuProLeuAlaPheThrCysSerGlnValSerThrThrSer 1044
3508CAGCGCGACCTGATCAGCCGTTA 3530
1045 ValSerGlnThrProSerAsnThrValLeuThrAsnThrIleMetSer 1061
3531 TCCCAATAGCGGTTTGAAGTATTTCCGACGCTCAACAGCGTTTGG 3580
1061 rSerAsnValThrSerThrAsnGlnLysProSerSerThrThr...Ser 1076
3581 CCGTACAGGAGCAATTAACCGCGCTATTCGCAAGACGCGCAACGCG 3630
1077 ProTrpAsnPheSer...SerGlyThrSerLeuProSerSerThrThr 1092
3631 GTTGGACAAACGCGCATCCGAGACCAAAACCTACCGTCCCAAGATT 3680
1092 o...SerGlnThrSerLeuSerThrAlaThrThrThrIleAsnGlyLeu 1108
3681 CC.....GCCTTACCGCGCAACAAACGACGCTGCGC 3712
1108 yStrThrValThrThrThrThrTrpCysProLeuAlaGlnLysSerThrValAla 1124
3713AAATCGTATGCAAGAAACCTGCGACGCGG 3744
1125 AlaSerSerGlnSerSerArgSerValAspArgPheValSerSer 3745
3745 CGCGTGGGATCCTGTTTCGACACCGGACCGAAACCTTCGAGCA 3794
1141 sProSerSerSerLeuSerGlnThrSerThrGlnThrThrLeuSer...T 1157
3795 CGGATCGGCAACTCGGACAGCGCTTGGCCACG.....GCG 3829
1157 hAlaThrThrThrThrThrSerGlyLeuThrValThrThrThrTrpCys 1173
3830 CCGTTTTCGCGCAATACG.....GCATCGACAGCTTCATCGGCAC 3873
1174 ProLeuThrSerThrSerThrLeuGlnAlaThrThrGlnThrSerThr 1190
3874 AGCGCGGCGCGGCTTTAGCAGCGCA.....G 3902
1190 rAlaLysValArgThrThrSerAlaSerSerAlaThrSerThrSer 1207
3903 CTTTCACAGCGCATCGGAGCAAAATCCGCGCGC..... 3937
1207 eLeuSerThrThrThrThrThrThrThrThrThrThrThrThrThr 1223
3938 ..CGGTCTGCATTAACGCATTCAGCAGCAT.....AC 3969
1224 GlnValCysSerGlyThrGlnCysThrGlnAspValProThrGlnSer 1240
3970 CGCGCGCGGTTTCG.....CGGATTCGCGCATCGACGCGCAT 4007
1240 rSerProAlaSerThrLeuAlaThrSerProSerValSerThrSer 1257
4008 CGCGCGACCGCTATTCGTCGCAAAAGCGGATTAACGTCAGCAAG 4057
1257 eTrSerPheSerThrThrThrAlaSerThrThrThrThrThrThr 1273
4058 TCAATATCCGCAACCCCGCGCTTCGATTCACACCGCTACCGCGG 4107
1274 SerVal.....ProLeuLeuProSerSerSerSerIleSerAla 1287
4108 AAGGAGCATTAATTCATAACCGCGGACACATTCATTCACACGCGCT 4157
1287 rSerProSerSerThrSerLeuLeuSerThrSerLeuProSerProAla 1304
4158 TTTGAGCCTGTCTTATACG..... 4177
1304 heThrSerSerThrLeuProThrAlaThrAlaValSerSerThrThr 1320
4178ATGCGCTTCGCGCAAGTCCGACACGCGCAATTC 4215
1321 lLeaSerSerLeuProLeuSerSerLysSerSerSerLeuSerThr 1337

4216 G 4216
1337 O 1337

seq_name: SwissProt_40:HLES_DROME

```

seq_documentation_block:
ID HLES_DROME STANDARD; PRT; 1077 AA.
AC Q02308;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Hairless protein.
GN H.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92387549; PubMed=1516831;
RA Bang A.G., Posakony J.W.;
RT "The Drosophila gene Hairless encodes a novel basic protein that
RT controls alternative cell fates in adult sensory organ development."
RL Genes Dev. 6:1752-1769(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93041287; PubMed=1419850;
RA Walter D., Stumm G., Kuhn K., Preiss A.;
RT "Hairless, a Drosophila gene involved in neural development, encodes
RT a novel, serine-rich protein."
RL Mech. Dev. 38:143-156(1992).
CC -1- FUNCTION: IS A POTENT ANTAGONIST OF NEUROGENIC GENE ACTIVITY
CC DURING SENSORY ORGAN DEVELOPMENT. THE EXPRESSION OF DISTINCT CELL
CC FATES BY THE TRICHOGEN (SHAFT) / TORMOGEN (SOCKET) SISTER CELL
CC PAIR DEPENDS ON THE LEVEL OF H ACTIVITY. A CERTAIN THRESHOLD LEVEL
CC OF H ACTIVITY IS REQUIRED, BELOW WHICH BOTH SISTER CELLS ADOPT THE
CC TORMOGEN FATE.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- TISSUE SPECIFICITY: OVARY, EMBRYOS, LARVAL AND PUPAL IMAGINAL
CC DISCS.
CC -1- SIMILARITY: CONTAINS A "PRD MOTIF".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M95192; AAA28607.1; ALT_INIT.
DR EMBL: X67239; CAA47664.1; -.
DR HSP: P04002; IWP.
DR Flybase: FBgn0001169; H.
KW Developmental protein; Nuclear protein; DNA-binding.
FT DOMAIN 115 123 NTR-RICH.
FT DOMAIN 642 648 POLY-SER.
FT DOMAIN 879 891 POLY-ALA.
FT DOMAIN 937 946 POLY-ALA.
FT DOMAIN 964 974 ALA-RICH.
FT DOMAIN 979 1008 HIS/PRO-RICH (PRD MOTIF).
FT CONFLICT 151 151 S -> A (IN REF. 2).
FT CONFLICT 702 703 OH -> L (IN REF. 2).
FT CONFLICT 891 891 A -> R (IN REF. 2).
FT CONFLICT 964 967 AAVA -> RLIP (IN REF. 2).
FT CONFLICT 974 974 MISSING (IN REF. 2).
FT SEQUENCE 1077 AA; 111039 MW; A94BFA27579E2F1 CRC64;

```

alignment_scores:

Quality: 211.00 Length: 1206
Ratio: 0.373 Gaps: 54
Percent Similarity: 46.849 Percent Identity: 20.398

alignment_block:
US-09-303-518D-649 x HLES_DROME ..

Align seg 1/1 to: HLES_DROME from: 1 to: 1077

```

979 AATGGGAATATCTTTTACGACGATATATATGGCAGGAAATCAAA 1028
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
30 AsnSerSerHisSerSerAsnAsnAsnAsnGlySerSerSerAsn 46
1029 TGGCAACATGACACACAACTCTCGCTTAATGATTAACACAGAC 1078
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
46 nAspAsnSerAsnAsnAspAlaSerSerSerAsnSerLysAsn. 62
1079 TTCAATTTGTTAATGTTCTTTATCCGACACGCAAGAACCTGTTAT 1128
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
63 .....AsnAsnThrSerAsnGlySerSerHis 71
1129 CATGCTGAGGTGTGTACACAGTTATCGACCCAGCATGATATGACGA 1178
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
72 .....SerAsnAsnAs 75
1179 AATATTTCTTTATGACGAGAAAGCGAATTGATATACACGCA 1228
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
75 nhrSerSerIleIleAlaGluAlaAlaAlaLysPheLeuLeuLysAsnG 92
1229 ACATCATCAAGTCCTCGAGATTTATTTCCAGAGATTTTACGGTC 1278
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
92 LysLeuAsn..... 94
1279 TCGCCTGAATATACGAACCTTGGCAGCGCGGCTTCATATTCAGTGA 1328
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
95 .....GlySerSer.SerThrs 100
1329 AGACAGTACCGTTA.....CTTGAAGATAAAGC 1357
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
100 eLyrProLeuProProProLeuProAlaAsnLeuSerArgThr 116
1358 GCGGCAACGACGCGCTGCCAAATCGCAAGGACCGCTGACG 1405
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 ThrProThrThrThrThrProSerSerSerSerSerThrAlaSerAs 133
1406 ....TTCAACCAAGGGAACCAAGGCTGATCAGC..... 1441
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
133 nGlyPheLeuProHisAlaLysThrProLysSerSerSerIleMetAla 150
1442 .....TGGCGACGCTACATGATTTTGTATGATGACGAGCAG 1477
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
150 lAserAlaAlaAlaAlaAlaSerValValValAlaThrAlaSerLysPro 166
1478 ACGATTAAGCAAAAAACAGCCCTTGTGTAATCGCTGTGTCAGCGC 1527
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 ThrIleAspValLeuGlyGlyValLeuAspLysSerSerLeuGlyVal 183
1528 AGGGGTACGTTGCACTGATCCGATTAATCAGTTCAACCCGCAAACT 1577
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
183 alaIthr.....GlySerLeuProThrThra 192
1578 CTATTTGCGCTTTCGCGCGC..... 1597
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
192 lAlaValAlaAlaAlaAlaLysThrAlaLysIleGlyLysGlySerAsn 208
1597 ..... 1597
209 SerClyGlySerPheAspMetLysArgThrProIleSerThrHisGlyAs 225
1598 .GACGTTTGATTAAGGCGATTCGC.....TTTGGTTCACCG 1637
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
225 nAsnSerTyrGlyTyrGlyGlyArgLeuGlnPhePheLysAspLyl 242

```



```

1638 TATTCAAAATACCGATG.....AGGGCGCATGA 1666
1642 yspheileleugluleuAlaArgSerLysAspLysAspLysSerLyr 258
1667 TTGTCACCAACATCAAGCAAGAAATCCACCGTTACCATTTACAGCAAT 1716
259 ValSerValThrArgLysThrPheArgProProSerAlaAlaThrSerAl 275
1717 AAAGATATTCGCTACACCGCAATTAACACCTTGATACGAAAGA 1766
275 aThrValThrProThrSerAlaValThrThrAlaThr...ProLysAsn 291
1767 AATTGGCTACACGGTTGGTTGGGAGAAAGATACGACCAAAACGACG 1816
291 LysAsnSerThrSerLeuSerPheSer.....Asp 300
1817 GGGCGCTACACCTGTTTACGAGCGCGCGGAGAGACCGACCGCTGCTG 1866
301 AspAsnSerSerLeuSerLeuSerProThrPheLysArgAspLysPro... 315
1867 CTTCGCGGAGACAAATTTAAACGCAACATCAGCAACAAACGAGCA 1916
316 ..TrpLysLysSerArg...ProArgArgLysLysSerLysLysLys 330
1917 ACTGTTTTCAGCGGACACCAACACCGGCTACATCATTTAAACG 1966
331 LeuPhePheHisArgProArgAsnSerThrLeuGlyArgAlaLeuLeu 347
1967 ACCATGCTCGCAAAAGAGCGCATTCCTCGCGGAAATCGGTTGGAGC 2016
347 gThrAlaAlaArgLysArgArgArg...ProHisGlyProLeuThrThrSer 363
2017 AAGCACTGG.....ATCAACCGCACTTTAAAGCGAAAC..... 2052
364 GluAspGlnGlnProIlePheAlaThrAlaIleLysAlaGluAsnGlyAs 380
2053 .....TTCCAATTAAAGCGGACAGCGGCTGTTCCCGCATATGTTG 2095
380 PASThrLeuLysAlaGluAlaAlaGluAlaGluIleGluAsnValAla 397
2096 CCAAGTGAAGAGCGATGGCATTTGAGCAATCAGCCCAAGCACTTTT 2145
397 LalaAlaAsp.....ThrThrThrAsn 404
2146 GGTGTCGACCGCATCAAGCAACATCTGTACACGTTGCGATGGAC 2195-
405 GluIleLysIleGluLysProAspThrIleLysGlyLysAspAlaGlu 421
2196 GGGTCTGACAAATGTGTGAAAGAAACATTAACGAGATTAAGTATG 2245
421 uArgLeuGluLysGluProLysLysAlaValSerAspAspSerGluSerL 438
2246 CTTCATTTGACT.....AAG 2259
438 ysgLysLysLeuProGlnGlnGlnValGluProGlnProLysAspGluThr 454
2260 ACCGACATCAGCGGCAATGCGATTCGGCATTCAGCTCATTTAAATCT 2309
455 ValAspValGluMetLysMetAsnThrSerGluAspGlnLys...Pro 470
2310 CACAGCGCTGCTG.....ACATCAACGCGCAATCTAGTACAA 2347
470 tThrGluLeuProArgIleThrAsnAlaValAsnGlyAspLeu.....A 485
2348 ATGGCGATACAGCTTATACAGTACAGCGCAACAGCCCAAC..... 2388
485 snGlyAspLeuLysAlaSerIleGlyLysProLysSerLysProLysPro 501
2389 AAGCGCAACCTTACG.....CTCGTGCGCAATCGCCCAAGC 2423
502 LysAlaLysLeuSerSerIleIleGlnLysLeuIleAspSerValProAl 518
2424 AACATTTAATCAAGCAGATTAAGCGCAACATCGGCTTCGGCGCAATG 2473

```

```

518 aArgLeuGluGlnMetSer.....LysThrSer..... 527
2474 CTTCATTTAATCTAAGCAGCAGCGCTACAAAGCGAGTGTACGCTT 2523
528 .....AlaValIleAlaSerThrThrThrSer 536
2524 TCCGGAACGCTAAGCAAGCAAGTAAAGCAATTCGCGCATAGGTAATGT 2573
537 SerAspArgIleGlyGlyLysLeuSerHis...AlaLeuThrHisLysVa 552
2574 CTCCCTGACGAGTAAAGCAGATTCATTTTGAAGCAGCGCTTTACCG 2623
552 LysProProSerSerAlaThr..... 559
2624 GACAAATACCGCGGCGCAAGATACGCGATTCATTAAGACAGCGAA 2673
560 .....AlaAlaGlyArgLeuValGluThrHisThrGlnHisValSer 573
2674 TGACGCGTGGCTACGACGCAAGTAAAGCAATTTAAACCTTGACAGCG 2723
574 ProArgLysArgIleLeuArgGluPheGluLysValSerLeuGluAsp... 589
2724 CACCATTAACATCAATTCGCGCTATCGCACAGATGCGGAGGCGGCA 2773
590 .....AsnGlyCysValAsnAsnGlySerGlyValAlaSerS 602
2774 CGGAGCTGCGACACATGCGCGCGCGCGCTTCGCGCGCTTCGCGCGCT 2823
602 erGlyGlyAla.....GlyGlyLysArgSerArgAla 612
2824 TCCCTATTATCCGTTTACACCGCACTTCGCTAGATACCGCTTTCACAC 2873
613 LysGlyThrSerThrSerProAlaGlyLysAlaSerProMetAsnLe 629
2874 GCTGACGTTAAACGCAATTAAGCAAGTGTACGAGCAATTCGCTTATGT 2923
629 uAlaProProGlnIleLysProSerProSer...ProGlySerSerSer 645
2924 CGGAGCTTTCGCTACCGGACGAGCA.....AATTGAAGCTGGCGGAA 2967
646 SerThrSerProAlaThrLeuSerThrGlnProThr...ArgLeuAsnSerS 662
2968 AGTTCGGAAGCACTTACCTTCGCGGCTACACATACCGGCAACGAC 3017
662 erTySerIleHisSerLeuLeuGlyLysSer..... 672
3018 TCGAAGCTGCAACATTAAGCGTACGTGAGAAAGAAACAAACAAACGC 3067
673 .....SerGlySerGlySerSerSerSerSerSer 683
3068 TGTCGAAACCTTAATTCACCGCTTCGCAAAACGACGTCGATCGCGC 3117
683 rGlyLysLysCysGlyAspHisProAla.....AlaI 694
3118 GCGTGCAGTTTACCACTATCCGCAAGACGCGAGTTCGCGCTGATAA 3167
694 LeIleSerAsnValHisHisProGlnHisSerMetTyGlnProSer... 709
3168 TCCGTCGAAAGAAACAAAGAGCTTCGCAACATTCGCGAA..... 3206
710 SerSerSerTyProArgAlaLeuLeuThrSerProLysSerProAspVa 726
3207 .....GGCAGAACGCAAAAGAGGCGGAGAA 3234
726 LserGlySerAsnGlyLysGlyLysSerProSerHisThrGlyThrL 743
3235 GACAAACGCGCAAGCTTGACGCGCT.....GATTGCGCGCGG 3272
743 yslArgSerProProTySerAlaGlySerProValAspTyGlnHis 759
3273 GCGGAGTCCGCTCGAAAGACAGAAACGTTGCGCAACGCGCGCGGAGG 3322

```

760 SerpHeTyArgAspPro..... 765
 3323 CAGGGGGGAAATGTGGCATTTATGAGCGGAGAGAGAAAAACGG 3372
 766 TyrLagLylalglYArgProSerThrs 775
 3373 GTGACAGCGGATTAAGACACCGCTTGGGGAAGACCGGAGGAGAAC 3422
 775 erLysSerAlaSerGlnAspLysSerProArgSerSerProAlaSer 791
 3423 CCGCGCGGCTACACACCGCTTCCCGG.....CG 3451
 792 ProAlaThrThrProArgThrValProLysLysThrAlaSerLleArgAr 808
 3452 CCGCGCGGCGCGCGCGGATTTGGCGCAACGACCCGACCGGACGCC 3501
 808 gLlurPheAlaSerProSerAlaSerSerSerSerSerProSerProGlyA 825
 3502 CAACCGGAGCGGACCTGATCGACCGCTTATGCCATAGCGTTGAGTGA 3551
 825 sPArgSerAlaSerProProGlnArg..... 833
 3552 ATTTTCGCGCAGCTCAACAGCT.....TTTGGCGGTACAGAGCAAT 3595
 834ArgHisMetGlnGlnProHisLeuGlnArgSerSerProle 848
 3596 TAGACCGCGTATTGGCGAAGACCGCGGCAACCGCTTGGACAGCGGC 3645
 848 whiStyTyrMetTyrProProProProGlnValAsn..GlyAsnGlyse 864
 3646 ATCCGGGACACCAACACTACCGCTTGGCAAGATTTCCGCGCTTACCGCA 3695
 864 rAlaGlySerProThr.....SerAlaPro..... 872
 3696 ACAAAACGACCTGCGCAATGCGTATGAGAAAAACCTCGGACGAGCGC 3745
 873ProThrSerAsnSerSerAlaAlaAlaValAlaAlaAlaAla 887
 3746 GCGTGGCATCTGTTTTCGCAACCGGACCGGCAACACCT..TCGAC 3792
 888 AlAlaAlaAlaAlaLylLleProSerProSerLylLleAsnProTyrLleSe 904
 3793 GAGCGCATGCGGCACTCGGCGGACGCG..... 3817
 904 rThlLeuAlaAlaLeuAlaArgHisAsnProLeuThrMetHisTyrGlnT 921
 3818TTGCCACGCGCGCGCTTTCGCGGCAATACGCGCA 3850
 921 hrGlyAlaSerProLeuLysSerProHisProGlnProGlyLysSerAla 937
 3851 TCGACAGGTTCTACATCGGCAATACGCGCG..... 3880
 938 AlAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 954
 3881GGCGGCGTTTGG 3893
 954 rHisAlaPheAlaLysAsnGlyAlaGlyAlaAlaAlaAlaAlaAlaAla 971
 3894 CAGCGGCGACCTTCAGACG..... 3914
 971 lAlAlaAlaAla..PheGlyGlnProAlaProSerProHisThrHisProH 987
 3915CATCGGAGGCAAAATCCGCGCGCTGTCGATTCAGGCAAT 3956
 987 sLeuAlaHisProHisGlnHisProHisProAlaAlaLeuThrThrHis 1004
 3957 C.....AGGCAATACCGCGCGCTTTCGCGGATTCGATTCGCAAC 4000
 1004 lSerProAlaHisLeuAlaThrProLysLeuThrAspSerSerThrAsp 1020
 4001 CGGACATCGCGGCAACGCGCTATTTCGTCAGAAAAAGCGATTCACGCTAC 4050
 1021 GlnMetSerAla.....ThrSerHisArgThrAlaSerTh 1033

4051 GAAACGTCAATATCCGACCCCGCGCTTGCATTCACCGCTACGCGC 4100
 1033 rSerProSerSerSerSerAlaSerAlaSerSerSerAlaAlaThrSerG 1050
 4101 GGGCA 4105
 1050 LylAla 1051
 seq_name: SwissProt_40:OMP_RICJA

seq_documentation_block:
 ID OMP_RICJA STANDARD; PRT; 1656 AA.
 AC 006553;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)
 DE (Surface protein antigen) (Cell surface antigen 5) (Scas5) (rOMPb)
 DE (rOMP B) (Contains: 120 kDa surface-exposed protein (surface protein antigen) (120 kDa outer membrane protein ompb); 32 kDa beta peptide).
 DE OMPB.
 GN OMPB.
 OS Rickettsia japonica.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiidae; Rickettsia.
 OX NCBI_TaxID=35790;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-YH;
 RA Uchiyama T.;
 RT "Sequencing of the gene encoding the protein rOMP B of Rickettsia japonica";
 RL Submitted (MAY-1997) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY SIMILARITY).
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-LAYER WITH HEXAGONAL SYMMETRY.
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAL OMPA/OMPB FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AB003681; BAA20138.1;
 DR InterPro: IPR003858; rompa_rOMPb.
 DR Pfam: PF02708; rompa_rOMPb; 1.
 KW Antigen; S-layer; Cell wall.
 FT CHAIN 1 1338 120 KDA SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1339 1656 32 KDA BETA PEPTIDE.
 FT DOMAIN 528 533 POLY-GLY.
 SQ SEQUENCE 1656 AA; 168097 MW; 3132A6C9D9599F CRC64;

alignment_scores:
 Quality: 210.50 Length: 1479
 Ratio: 0.305 Gaps: 73
 Percent Similarity: 46.721 Percent Identity: 19.675

alignment_block:
 US-09-303-518D-649 x OMPB_RICJA ..

Align seg 1/1 to: OMPB_RICJA from: 1 to: 1656

661 GCAGTGCCTATTCTTGCTGCTGTTGCGCAATACCTTTCGCAAAATGG 710
 |||||
 |||||

```
251 AlaSnaIleuAsnIeuGlnAlaGlyThrIleAsnPheAsnGI 267
711 ATCAGTGTG...GCCACAGTCACTTAGTAGTAAAAAATTAACATA 757
::: ||| ::| ||| ::| ::|
267 yThrAspGlyThrGlyArgLeuValIleuSerIysAsnGlyAlaIat 284
758 GCCCATATGGTTTTCACACAGAGAGGCTCATTTGGCGACAGTGCCTCA 807
::: ::| ::| ::| ::| ::| ::| ::| ::|
284 hrAspPheAsn...ValThrGlySerLeuGlyCylAsnLeuIys 297
808 CCAATGTTATCTATGATGCCCAAAAGCAAAAGTGAATTAATGAGGCT 857
::: ::| ::| ::| ::| ::| ::| ::| ::|
298 GlyIleIle...GluLeuSnrThrValAlaIleAsnGlyCyl 310
858 ATTG...CAACGCGCAACCCCTATATAGCAAAAGCAATG 895
||| ::| ::| ::| ::| ::| ::| ::|
310 nleuIleAlaAsnAlaGlyProAlaAsnAlaValIleGlyThrAsnSng 327
896 GCTCCAGTGTGCTAAAGATTGCTATGATGAATCTTGTGCA 945
||| ::| ::| ::| ::| ::| ::| ::|
327 yAlaGlyArgAla...AlaGly 333
946 GATACCCATTGATTTCTACAGACACAGCAAAATGGAAATCTCTT 995
::: ||||| ::| ::| ::| ::| ::| ::|
334 PheValIleSerVal...AspAsnGlyLysAlaAlaIatH 345
996 TAACGACGATAAT...AATG 1012
||| ::| ::| ::| ::| ::| ::|
345 rIleAspGlyGlnValIleAlaIleAlaIleGlnSerAlaAsn 362
1013 GCACAGCAAAATCAATGCCAAACATGACACAAATCTCTGCTAATAGA 1062
::: ||||| ::| ::| ::| ::| ::| ::|
362 IaAsnGlyGlnValIleAsnPheArgHis... 370
1063 TTAAACACGCAACCGTTCAATGTTATGTTCTTTATCCGAGACAGC 1112
::: ||||| ::| ::| ::| ::| ::| ::|
371 .....IleValAspValGlyIleAspGlyThrH 380
1113 AAGAGAACCTGTTATCATGCTGACAGTGTGTCACACACTTATCCACCA 1162
::: ||||| ::| ::| ::| ::| ::| ::|
380 r...AlaPheIleThrAlaIleSerIleValAla... 390
1163 GACTGATATATGAGAAATATTTCTTTATTCAGCAAGCAAA... 1206
::: ||||| ::| ::| ::| ::| ::| ::|
391 ...IleThrGlnAsnSerAsnPheGlyThrThrAspPheGlyAsnLeuAla 406
1207 .....GCCGATGATCTTACCAGACATC... 1233
407 AlaGlnValIleValIleProAspThrIleThrLeuThrGlyAsnPheThrGI 423
1234 .....AATCAAGT...GCTGAGAGATTATTTCCAAAGAG 1267
||| ||| ||||| ::| ::| ::| ::|
423 yAspAlaAsnAsnProGlyAsnThrAlaGlyAlaIleThrPheAlaAla 440
1268 ATTTTACGGCTCGCCTGAAATATACGAAACTTGCAAGCGCGCGCTT 1317
::: ||||| ::| ::| ::| ::| ::| ::|
440 sngIyThrIleuAlaSerAlaSerAlaAsp...AlaAsnVal 452
1318 CATATCAGTGAAGACATGACGTTACTTGAAGTAACGCGCGGCAAA 1367
::: ||||| ::| ::| ::| ::| ::| ::|
453 AlaValIleThrAsnAsnIleThrAlaIle...GluAlaSerGlyValGlyVa 468
1368 CGACCGCTGTCCAAATTCGCAAGGACGCTGACCTTCAAGCCCAAG 1417
::: ||||| ::| ::| ::| ::| ::| ::|
468 lValGlnIleuSer...GlyThrIleThrAlaGlyIleuArgL 481
1418 GGGAAACCAAGCTCGATC...AGCGTGGCGACGCTACAGCACTTTG 1464
||| ||||| ::| ::| ::| ::| ::| ::|
481 euGlyAsnIleGlySerValIlePheIleuAlaIleGlyThrValIle... 496
1465 GATCAGCAGCGACGATTAAGCAAAACCAAGCCTTATGTAATCGG 1514
::: ||||| ::| ::| ::| ::| ::| ::|
497 .....AsnGlyLysValIleAsnGlnThrValIleValGI 507
1515 CTGGTACGCGAGGCTACGCTGCACTGAATGCCATATATCAGTTCA 1564
||| ::| ::| ::| ::| ::| ::| ::|
507 yGlyValIleuAlaIleGlyAlaIleThrIleuAspIleSerAlaThrIle 524
1565 ACCCGCAACAACCTCTTTGGCGCTTCCGCGGAGCGTGTGATTAAC 1614
::: ||| ::| ::| ::| ::| ::| ::|
524 hrGlyAsp...IleGlyAsnGlyGlyGlyAlaIleuGln 537
1615 GGGCATTCCTTCGTTCCACCGTATTCAAAT...ACCGATGAAGGCGC 1661
::: ||||| ::| ::| ::| ::| ::| ::|
538 SerIleThrIleuAlaAsnAspAlaThrIleThrIleuThrIleuGlyAla 554
1662 GATGATTCACACCAAT...CAACGCAAGAAT 1693
||| ::| ::| ::| ::| ::| ::|
554 AsnIleIleSerAlaAsnGlyGlyThrIleAsnPheGlnAlaIleuGln 571
1694 CCACGCTTACCATTCACGCAATAAAGAT... 1722
::: ||||| ::| ::| ::| ::| ::| ::|
571 yThrIleIleuThrIleuThrIleuThrIleuThrIleuThrIleuThrIleu 587
1723 .....ATTGCTACA...ACCGC...AATACACAGCTTGA 1754
||| ||||| ||||| ::| ::| ::| ::| ::|
588 LeuAlaIleAlaIleThrAspGlnThrGlyValIleAspIleSerIleuThr 604
1755 TAGCAAAAAGAAATTCCTACACAGGTTGTTGGCGAGAAAGATACGA 1804
::: ::| ::| ::| ::| ::| ::| ::|
604 rAsnAlaGlnThrIleuThrIleuSerGlyThrIleGlyIleIleGlyAla 621
1805 CCAAAACGAAAC...GGCGGCTCAACCTGTTTACAGACGCGCGGAGAA 1851
::: ||||| ::| ::| ::| ::| ::| ::|
621 sAsnIleThrIleuGlyGlnPheAsnIle...GlySer 632
1852 GACCGCACCTGCTGCTTCCGCGGCAAAATTAAGCGCAACATCAC 1901
::: ||||| ||| ::| ::| ::| ::| ::| ::|
633 SerIleThrIleuAsnGlyIleAsnValAlaIleAsnGlnIleuValIle 649
1902 GCAAAACAAGCGCAACCTGTTTCAAGCGGACAGCCACCGCGCT 1951
::: ||||| ::| ::| ::| ::| ::| ::|
649 eGlyAsnIleIleSerValIleGlnPhe...AlaH 659
1952 ACAAT...CATTAACGACCATGCTGCGCAAAAGAGCGCATTCCTG 1998
||| ::| ::| ::| ::| ::| ::|
659 IsAsnThrIleuIleThrIleuThrIleuThrIleuThrIleuThrIleu 673
1999 GGGGAATGCTGTGGGACACAGCTGATCAACCGCATTTAAAGCGA 2048
||| ::| ::| ::| ::| ::| ::|
674 GlyLysIleIlePheAsn... 679
2049 AAACCTCCAAATTAAGGCGGACAGCGGCTTCCCGCAATGTGCA 2098
||| ||||| ::| ::| ::| ::| ::|
680 .....ProValIleAsnAsnIleThrIle 688
2099 AAGTGAAGCGCATTTGACATTCAGCCCAAGCAAGT... 2142
||| ::| ::| ::| ::| ::| ::|
688 euAlaIleGlyThr...AsnLeuGlySerAlaIleAsnProIleuAlaGln 703
2143 .....TTGGTGTGCGACCGCATCAAGCGACCAATCTGTACACGCTG 2186
||||| ::| ::| ::| ::| ::| ::|
704 IleAsnPheGlySerLysGlyAlaArgAlaAspThrValIleuAsnValGI 720
2187 GGACTGACGGGCTG...ACAAATGTGTGCAAAACCAATTAAC 2230
::: ||||| ||||| |||||
720 yGlnGlyValAsnLeuIleuIleuIleuIleuIleuIleuIleuIleuIleu 734
2231 ACAGTAAAGTG...ATTGCT 2247
||| ::| ::| ::| ::| ::| ::|
734 sPalaAsnValIleGlySerPheValIleAsnAlaGlyGlyLysAsnIleVal 750
2248 TCATGTACTAAGACGACATCAGCGCAAT...GTGATGCT 2285
||| ||| ::| ::| ::| ::| ::|
751 SerGlyThrValIleGlyGlnGlnIleuLysPheAsnThrValAlaIle 767
```



```

1312 aLThrSPfrrSerAlaAlaIleThrAspAlaGlnLeuAspValAlaLys 1322
3832 ... GTTTGGGGAATACGGCATCGACAGTTC.....TACAT 3866
1329 AspIleGlnIleAlaGlnLeuIleu...AsnArgLeuGlyAlaLeuArgTyle 1344a
3867 CGGCATACACGGCGGGCGGTTTATGACAGCGGAGCCCTTCACAGCGCA 3916
1344 uGlyThrProGluMetValGlySerGlnIleAlaIleProAlaIalav 1361
3917 TC...GGAGCGAAATTCGCGCGCGCGCTGTCGATTACGGATTACAGCA 3963
1361 aLAlaAlaGlyAspGluIuAlaValAspAsnValAlaTyrGlyIleThrPala 1377
3964 CGA.....TACGCGCGCGGTTTCGCGG 3986
1378 LysProPheTyrThrAspAlaHisGlnSerLysGlyGlyLeuAlaG 1394
3987 ATTCCGC..... 3993
1394 YThrLysAlaLysThrThrGlyIleAlaIleGlyLeuAspThrLeuAla 1411
3994 .....ATCGAACCGCACATCGCGCACAGCGCTATTTCGTC 4029
1411 snAsnAsnLeuMetIleGlyAlaAlaIleGlyIleThrLysThrAspIle 1427
4030 CAAAAAGCGATATCCCG.....TAGAAACGTCAAATATCGCCAC 4070
1428 LysHisGlnAspTyrLysLysGlyAspLysThrAspValAsn..... 1441
4071 CCGCGCGCTTCGATTCAACGCGTACCGCGG..... 4101
1442 ...GlyPheSerPheSerLeuTyrGlyGlnGlnPheValGlnAsp 1457
4101 ..... 4101
1457 hepAlaGlnGlySerAlaIlePheSerLeuAsnGlnValLysAsnLys 1473
4101 ..... 4101
1474 SerGlnArgTyrPhePheAspAlaAsnGlyAsnMetSerLysGlnIleAl 1490
4102 .....GGCATTAAGGACAGATT 4117
1490 aAlaGlyAsnTyrAspAsnMetThrPheGlyLysAsnLeuThrValGlyT 1507
4118 ATTCAATTCAACCGCGCAACACATTTCCATCAGCGCTATTGAGCGTCG 4167
1507 yraspTyrAsnAlaMetGlnGlyValLeuValThrPheMetAlaIleu 1523
4168 TCGTATACCGATGCGCGCTTCGGGCAAAAGTCGAGACAGCGTCAATACGC 4217
1524 SetTyrLeuLysSerSerAspLysAsnTyrLysGluThrGlyThrThva 1540
4218 CGTATTGGCTCAGGATTTGGGCAAA...ACCGCAGTCGGAATGGGCG 4264
1540 lAlaAsnLysGlnValAsnSerLysPheSerAspArgThrAspLeuIle 1557
4265 TAAAGCGCGAAATCAAGGTTTCACGCGTGCCTC 4299
1557 aGlyAlaLysValAlaGlyIleThrMetAsnIle 1568

seq_name: SwissProt_40:HLVA_SERMA

seq_documentation_block:
ID HLVA_SERMA STANDARD; PRT; 1608 AA.
AC P15320;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, last sequence update)
DT 01-NOV-1990 (Rel. 16, last annotation update)
DE Hemolysin precursor.
SHA.....
OS Serattia marcescens.

```

CC	Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC	Serratia.
OX	NCBI_Taxid=615;
RN	(1)
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.
RX	STRAIN-SN8;
RX	MEDLINE=88257037; Pubmed=3290200;
RA	Poole K., Schiebel E., Braun V.;
RT	"Molecular characterization of the hemolysin determinant of <i>Serratia</i>
RT	<i>marcescens</i> .";
RL	<i>J. Bacteriol.</i> 170:3177-3188(1988).
CC	-I- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD
CC	CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY
CC	DEFINED.
CC	-I- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM
CC	ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA
CC	REQUIRES SHLB FUNCTION.
CC	-I- SUBCELLULAR LOCATION: TO THE ERYTHROCYTE MEMBRANE. SHLA
CC	-I- SIMILARITY: TO P. MIRABILIS HEMOLYSIN (HPMA).
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC	or send an email to license@sdb-slb.ch).
CC	-----
DR	EMBL: M22618; AAA50323.1; -.
DR	PIR: A28182; A28182.
KW	Hemolysis; Toxin; Outer membrane; Signal.
FT	SIGNAL
FT	CHAIN 1 30
FT	31 1608 HEMOLYSIN.
SQ	SEQUENCE 1608 AA; 165078 MW; D6659BA76EE7DA0D51 CRC64;

alignment_scores:	
Quality:	210.00
Ratio:	0.324
Percent Similarity:	45.290
Length:	1433
Gaps:	67
Percent Identity:	19.819

Align seg 1/1 to: HLYA_SERMA from: 1 to: 1608
 US-09-303-518D-649 x HLYA_SERMA ..

```

649  TCAATCATATTCGAAGGCGCATTCATTGCGTCGTTGGTGCATACCTT 698
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
251  SerIyrrYrLeuNglySerMetGlnAlaIyArgIleasnIleIeaSnth 267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
699  TGCCAAATGATCAGT.....GGGCGACATGCATCTAGTA 739
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
267  rAlagIn...GlySerGlyAllyLysLeuAlaIySerLeuAsnAlaIyA 293
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
740  GTGAAAAATTAAACATAGCCCATGTGTTTTTACCACA..... 780
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
283  spGIn...LeuYsValIlyAlaIyrrAspIleArgSerGInSerIArgVal 298
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
781  .....GGAGGCGCATTTGGCCACACTGG 803
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
299  AspAspIaSerSerIasnIySasnGlyIyAspAsnIyGlnAsnIyArg 315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
804  CTCACCAATCTTTATCTGATGATGCCAAAAGCAAAAGTGGTTAATAATG 853
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
315  gGlyIyIleIyValAsnAspArgSerSerGInIrrLeuThrArgT 332
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
854  GGGATTTCGAAGGGCAACCCCTATATGAAAAAGCATAGCTTCCAG 903
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
332  hrcIleuIySgIyLysAsnIleSerLeuValAlaAspAsnIlaSnIs 348
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
904  CTGGTTTCGTAAGAT..... 918
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
349  LeuThrAlaIThrAspIleArgIyGlnAspIleThrIrrLeuGlnGlyIy 365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```
918 ..... 918
365 sleuthrleuaspglynglnleuylsglnthrglnghisthraspa 382
919 ..... TGGTCTATGATGAATCTTTCGTGAGATACCATTCAGTATTC 963
382 spargttrpheriysertrp ..... glntryasprvalthr 393
964 TACGAACCACTCAAAATGGGAATCTTTTAAACGACGATATATATGG 1013
394 ATGGLuArgGlnGlnleuGlnleuGlnleuGlnleuValalaalaserG1 410
1014 CACGGAATAATCAATGCCAAATGACAAATCTCTGCCATATGAT 1063
410 yserlalaYserlaleuIleSerthrglnGlu ..... 419
1064 TAAAAACAGAACCTTCATTTGTTTAAATGTTTCTTATCC ..... 1104
420 ..... AsprVallyleuLeuGlyAlaAsnValSerAlaAsprArg 432
1105 ..... GACACAGACAGAACCTGTTTATCATGCTGCGAGTGG 1142
433 AlaLeuSerVallylalaalargasp ..... ValHisLeuAlaGlyLe 447
1143 TGTC ..... AACAGTTATCGACCCAGACTGAATATGAGAGAA 1180
447 uValGlnLysAsprLysSerSerLysArgGlyTyrglnArgasn ..... H 462
1181 ATATTTCTTATTTATGACGAAGAAAGGCAATGTACTTACCAAC 1230
462 IsthrSerSerLeuArgthrglyArg ..... Trpserasn 473
1231 ATCAATCAAGT ..... GCTGAGATATATATTCGAAAGAGA 1268
474 SerAsprLysSerLysSerLysAlaSerGluLeuArgSerGlnGly 490
1269 TTTTACGCTCTCGCTTAAATACGAACCTTGGCAAGCGCGGCTGTC 1318
490 uleuthrleuYsAlaGlnArgasnValSerthrglnGlnGlyAlaValH 507
1319 ATATCAGTGAACAGTACCGTT ..... 1341
507 lAlaGlnArgAsprleuthrIleAsprAlaAsnGlnIleGlnValGly 523
1342 ..... ACTTGAAGTAAACGGCGTGCACAAACCGCTGCCAA 1382
524 ValGlnLysThrAlaAsnAlaLysAlaValArgAsprLysThrSerTr 540
1383 AATCGC ..... 1389
540 pGlyGlyIleGlyGlyAlaAsprAsnLysAsnSerAsnArgGly 557
1390 ..... AAAGCAGCTGCACGTT ..... 1407
557 lSerHisAlaSerGluLeuthrSerGlyThrIleuArgLeuAsnGly 573
1408 ..... CAAAGCAAGGCAAGCAAGCAAG 1430
574 GlnGlnGlyValThrIleThrGlySerLysAlaArgGlyGlnGly 590
1431 CTGATCAACCTGGCGAGCGTACAGTATTTGGATCAGACGACGAC 1479
590 yGlyValThrAlaThrHisGlyGlyLeuArgIleAsprAsnAlaLeuSerTr 607
1480 ..... GATTAAGCAAAACAAAGCTTTAGTGAATCGGCTGGTC 1521
607 hThrValAsprLysIleAsprAlaArgThrGlyThrAlaAsnIleThr 623
1522 AGCGGAGGAGGTACGCTGCACTGAATCCGATATACGTTTC ..... 1563
624 SerSer ..... SerHisLysAlaAsnSerTrGlnSerSe 636
1564 ..... AAC 1567
636 rThrAlaSerGluLeuLysSerAsprHisAsnLeuThrIleValSerHisL 653
1568 CCGAACAACTCTATTTGCGCTTCGC ..... GCCGAGCTTTGGAT 1608
653 yAsprAlaAsprValIleGlySerGlnValAlaSerGlyGlyGluLeuSer 669
1609 TTA ..... AACGGCATTCGTTTCGTCACCGTATTC 1643
669 ValGlnSerLysThrGlyAsnIleAsnValLysAlaAlaGlnArgGlnG1 686
1644 AATATCCGATGAA ..... GGGCGATGATTCAC ..... 1674
686 nasnIleAsprGlnGlnLysThrAlaLeuthrValAsnGlyTyrgAlaLysG 703
1675 ..... CCAATCAA 1683
703 LuAlaGlyAsprLysGlnTyrgAlaGlyLeuArgIleGlnHisThrArg 719
1684 GACAAAGATCCACCGTTACCATTTACAGCAATAAGATATTGCTACAC 1733
720 AsprSerGlnLysThr ..... ThrArgThrGlnAsnSerAlaSerSerLeuSe 735
1734 CGGCAATTAACAACCTTGATACGCAAAAGAAATGCCACAGCGT 1782
735 rGlyGlySerValLysLeuLysAlaGlnLysAsprValThrPheserGlyS 752
1783 ..... TGTTTGGCGAAGATACGCAAAACGACGCGGCTCAC 1827
752 erLysLeuValAlaAsprLysGlyAsprAlaSerValSerGly ..... Asn 766
1828 CTGTTTACCACCGCCGCGACAGACCGCACCTTCGCTTCGCGGG 1877
767 LysValSerPheluAlaAlaAsprLysThrAlaSerAsnThrGlnG1 783
1878 AACAAATTTAAAGCGCAATCACGCAACAACAGCGC ..... AACSTGT 1921
783 nThrLysIleGlyGlyPheTyrglyThrGlyLysAsprLysLeu 799
1922 TTTTACCGCGACAGCACACGCGACGCTTACATCATTTAAACGACAT 1971
800 ..... GlySerGlyValGlnAlaGlyTyrgLysAsnLysThrGln 813
1972 TGCTCGCAAAAGAGGCAATCTCTCGCGGAAATGCTGCGACACGA 2021
814 AlaGlnSerLysAlaIleThrSerGlySerAsprValLysGlyAsnLe 830
2022 CTGATCAACCGCACATTTAAAGCGCAAACTTCAAAATTTAAAGCGGAC 2071
830 uThrIleAsnAlaArgAsprLysLeuThrGln ..... GlnGlyAlaG 844
2072 AGCGGCTGGTT ..... TCCGCAATGCTGCCAAAGTGAAGGC 2109
844 lHisSerValGlyGlyAlaTyrgLysAlaAsnAlaAlaGlyValAsp 859
2110 GATTGCGCATTTG ..... AGCAATCACGC 2132
860 ..... HisLeuAlaAlaAlaAsprThrAlaSerThrThrThrLysThr 874
2133 CCAACGACTTTTGGTGTGCGACCGCATCAAGCGCAACATCTGACAC 2182
874 rAsprValGlyValAsnIleGlyAlaAsnValAsprLysSerAlaValThr 891
2183 GTTCG ..... GACTGACGCGGT 2199
891 rGpProValGlnuArgAlaValGlyLysAlaAlaLysLeuAsprAlaThrGly 907
2200 CTGACAAATTTGTC ..... 2214
908 ValIleAsnAsprIleGlyGlyLysGlyAlaProAsnValGlyLeuAsprI1 924
2215 ..... GAAAAAACCATTAACGACGATTAAGTGA 2242
```

```
924 eGlyIaGlnGlySerSerGlyLysArgSerSerSerGlnAlaVal 941
2243 TTGGTCACTTACTAAGACCGACATCAAGCGCAATGTCATGCTT 2286
941 aValSerSerValGln.....AlaGlySerIleAspIleAsnIle 954
2287 .....GCCGATCAGCTCA 2300
955 LysGlyGluValArgAspGlnGlyThrGlnTyrglnAlaSerLysGlyAl 971
2301 TTTAACTCACA.....GGCTTCACACACACACACGCA 2335
971 aValAsnLeuThrAlaAspSerHisArgSerGlnIleAlaIleAsnIle 988
2336 ATCTTACTGCAATGGCGATACAGT.....TAT 2364
988 InaSpGlnSerArgAspThrArgGlySerAlaGlyValArgValTyr 1004
2365 ACAGTCAGC.....CACACGGCAC 2384
1005 ThrThrThrGlySerAspLeuThrValAspAlaLysGlyGlnGlyThr 1021
2385 CCAAAACGGCAACTTACGCTCGGGCAATGCCAAGCAACATTTATC 2434
1021 rGlnArgSerAsnSerAla.....SerG 1030
2435 AACCCACATTTAAACGCAACATCGCTTCGGCAATGCTTCATTTAT 2484
1030 InAla...ValThrGlySerIleAspAlaIleAsnGlyIleAsnValAsn 1045
2485 CTAAACGACACGCGCTACAAAAGCGAGCTGACGCTTCGCGCAACG 2534
1046 ValLysLysAspAlaIleTyrGlnGlyThrAlaLeuAsnGlyGlyThr 1062
2535 TAAAGCAACGTA..... 2547
1062 YLysThrAlaValAsnAlaGlyLysPrlLeuArgLeuAspGlnAlaSer 1079
2548 .....AGCATTCGCGACACTACGGTAAATGTCCTCCAGCATTAAG 2589
1079 pRlyGlnSerGlnSerArgSerGlyPheAsnValLysAlaSerAlaLys 1095
2590 GCACTATTCATTTGAAACGACGCGCTT..... 2619
1096 GlyGlyPheThrAlaAspSerLysAsnPheGlyAlaGlyPheGlyGly 1112
2620 .....ACCGACAAATCA 2632
1112 YThrHisAsnGlyLysSerSerSerSerThrAlaGlnValGlyAsnIle 1129
2633 GCGGC.....GGCAAGATAGCGCATTTACAC 2658
1129 eGlyGlnGlnGlnGlyAlaGlyLeuLysAlaGlyArgAspLeuThrLeu 1144
2659 TTTAAAGACAGCAATGACGCTGCGGTACAGCAGCAATTA..... 2700
1145 .....GlnGlyThrAspValLysSerGln 1152
2701 .GGCAATTTAACTTGACAAAGCCACCATTAACAATTCGCGCATC 2749
1152 nGlyAspValSerLeuSerAlaGly.....AsnLysValAlaIle 1165
2750 GCCAGATGCGGCGGCGGCGGCGGAGTCGACAGATGGCGCGCC 2799
1165 eGlnAlaIleAlaGlnSerThrGlnThr.....Arg 1174
2800 CGCGCTTGGCGCGCTGCGCGCTTCCCTATTAATCCGTACACGGCAAC 2849
1175 LysGlnSerLysLeuSerGlyAsnIleAspLeuGlyAlaGlySerSerAs 1191
2850 TTCCGTAGAAATCCCGTTCAACAGCGTGAAGGTAACGGC..... 2889
```

```
1191 pSerLysGlnLysThrGlyLysAsnLeuSerAlaGlyAlaPheAspI 1208
2890 .....AAATGAACGCTCAGGGAACATTCGCTTATATGCGAAGCTTC 2934
1208 LeAlaLysValAsnGlnSerAlaThrGlnGlnGlnGlnAlaThrIle 1223
2935 GCGTACCCGACGACAAATTTGACGCTGCGGAAGTTCCGAGAGCACTTA 2984
1224 .....AlaSerAspGlyLysVal 1229
2985 CACCTTGGCGGTCAACAATACCGGCAAGCAAGCTCGAAGCTCGAACAAT 3034
1229 LThrLeuSerAlaAsnGlnLysGlyAspAla..... 1240
3035 TGACGGTAGTGGAAGAAAGACAAACACCGCTGTCGAAAACCTTAAT 3084
1240 ..... 1240
3085 TTCACCTGCAAAACGACACGTCGATGCGCGCGCTGCGCTTACCACT 3134
1241 LeuHisLeuGlnGlnLysValSerGlyLysSer.....AlaAlaLe 1255
3135 CATCCGCAAGACGCGAGTTCGCGCTGATATCGGTCACAAAGACAG 3184
1255 uGlnAlaLysAsnGlyGlyIleLeuLeuGlnSerAlaLysAsnGlnGln 1272
3185 AGCTTTCGAC...AACTCGC...AAGCAAGACCAAA.....AAA 3222
1272 LsLysAspAsnThrSerLeuGlnIleLysAlaAsnAlaLysGlyGln 1288
3223 CAGCGCAAAAGAC.....AA 3239
1289 ThrPheAsnLysAspAlaGlyLysValAspProAsnThrGlyLysAs 1305
3240 CGGCAAAAGCTTTCACGCGCTGATTCGCGCGCGGCGGATCGCGTGA 3289
1305 pThrHisThrLeuGlnGlyAlaGlyLeuLysValGlyGlnGlnAspI 1322
3290 AACAAGAAAGCGTTCCGACGCGCGCGCGGAGCGGCGGCAAAATGTC 3339
1322 YThrThrHisAla.....AsnThr 1328
3340 GGCATTAAGCAGCGCGGAGAAAGAAAGCGGTGACGCGGATTAAGA 3389
1329 GlyIleThrAlaGlyAsp.....ValThrLeuAsnSerGlyLysAs 1342
3390 CACGCGCTTGGCGAAACAGCGC...GAAGCGAAACCGCGCGCTCA 3436
1342 pThrArgLeuAlaGlyAlaArgValAspAlaAspSerValGlnGlyLys 1359
3437 CGGCTTCCCGCGCGCGCGCGCGCGCGCGGATTTGCGCAACTGCA 3486
1359 aGlyGlyAspLeuHisValGlnSerArgLysAspAla..... 1371
3487 CCCCACCGGAGCCCAACCGGACGCGCTGATCAGCGGTATAGCA 3536
1372 .....GlnAsnGlyValLysValAspValAs 1380
3537 TAGCGTTAGTAGAATTTCCGCCACGCTCAACAGCTTTCCGCGTAC 3586
1380 pAlaGlyLeuSerHisSerAsnAspProGlySer..... 1392
3587 AGAGCAATTTAGACCGCTATTTCCGAAGACGCGCGCAAGCGCTTGG 3636
1393 .....IleThrSerLysLeuSerValGlyThrProArgTyrAl 1406
3637 ACAAGCGATCCGGGACCAACACTACGTTCCGAAGATTTCCGCGC 3686
1406 aGlyLysValLysGlnLysLeuGlnAlaGlyValAsnLysValAlaAsp 1423
3687 CTACGCGCAACAAACGACTGCGCAAAATCGATGCAAAAA..... 3730
1423 LaThrThrAspLys.....TyrAsnSerValAlaArgArgLeuAsp 1436
```



```

3731 .....ACCTGGCAGCGGCGCTGGCAGCTGTTTCG 3765
1437 ProGlnGlnAspThrThrGlyAlaValSerPheSerLysAlaGlnGly 1453
3766 CACACCGGACCGAAGAACCT..... 3787
1453 sValThrLeuProAlaThrProAlaGlyLysProGlnGlyProLeu 1470
3788 .....TCGACGACG 3796
1470 rPaSpArgGlyAlaArgThrValGlyLysAlaValLysAspSerIleThr 1486
3797 GCATCGGCAACTGGCAGCGC...TGGCCAGCGCGCTTTCGGGCA 3843
1487 GlyProAlaGlyArgGlnGlnLysLeuValAsnAlaAspValValAs 1503
3844 TACGGCATCGACAGTTCTACATCGCATCGCGC.....GGCGCG 3886
1503 nasnasnAlaValGlyGlnGlnSerAlaIleAlaGlyLysAsnGlyVala 1520
3887 GTTTAGCAGCGGCGACCTTTCAGACGCGATCGAGCAAAATCCG 3933
1520 lalenglnvalGlyGlnThrGlnThrGlnGlyGlnIleArg 1535
seq_name: SwissProt_40: HXA2_HAEIN

```

```

seq_documentation_block:
ID HXA2_HAEIN STANDARD; PRT; 928 AA.

```

```

AC P4354;
DE 01-NOV-1995 (Rel. 32, Created)
DE 01-NOV-1995 (Rel. 32, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heme/hemopexin-binding protein precursor (Heme:hemopexin utilization
DE protein A).
GN HXA2.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 22-35.
RC STRAIN-DL42 / SEROTYPE B;
RX MEDLINE=95115556; PubMed=7815944;
RA Cope L.D., Thomas S.E., Latimer J.L., Slaughter C.A.,
RA Mueller-Eberhard U., Hansen E.J.;
RT "The 100 kDa heme:haemopexin-binding protein of Haemophilus
RT influenzae: structure and localization."
RL Mol. Microbiol. 13:863-873(1994).
RN [2]
RN SEQUENCE OF 1-30 FROM N.A.
RC STRAIN-DL42 / SEROTYPE B;
RX MEDLINE=95270579; PubMed=7751272;
RA Cope L.D., Yogev R., Mueller-Eberhard U., Hansen E.J.;
RT "A gene cluster involved in the utilization of both free heme and
RT heme:hemopexin by Haemophilus influenzae type b."
RL J. Bacteriol. 177:2644-2653(1995).
CC -1- FUNCTION: BINDS HEME/HEMOPEXIN COMPLEXES.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: U08348; AAA74138.1;
KW Transport; Signal; Repeat.
FT SIGNAL 1 21
FT CHAIN 22 928 HEME/HEMOPEXIN-BINDING PROTEIN.
FT DOMAIN 101 679 6 X 6 AA APPROXIMATE REPEATS.

```

```

FT REPEAT 101 106 1-1.
FT REPEAT 205 210 1-2.
FT REPEAT 279 284 1-3.
FT REPEAT 410 415 1-4.
FT REPEAT 635 640 1-5.
FT REPEAT 674 679 1-6.
FT DOMAIN 149 172 4 X 6 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 149 154 2-1.
FT REPEAT 155 160 2-2.
FT REPEAT 161 166 2-3.
FT REPEAT 167 172 2-4.
SO SEQUENCE 928 AA; 101228 MW; 67DA5466A4B92390 CRC64;

```

```

alignment_scores:
Quality: 208.50 Length: 1107
Ratio: 0.419 Gaps: 59
Percent Similarity: 44.986 Percent Identity: 19.874

```

```

alignment_block:
US-09-303-518D-649 x HXA2_HAEIN

```

```

Align seg 1/1 to: HXA2_HAEIN from: 1 to: 928

```

```

454 GCCATCCCTTATGCGCGCATATCATATGCGGTTGCATAATTTGT 503
17 GlyAlaAlaIleValAlaSerThrProAspPheProGlnHisHisLysThrVa 33
504 CACAGATCCAGACCTGTGAATGACCACTTATATGATGGCGGAAAT 553
33 lPheGlyThrValThrIleGlyLysThr..... 43
554 ATATGATCAAAATATATACCTGACCGTCTGCTATTTGGCGAGCAG 603
44 .....AlaAspLysMetThrIleLysGlnGlySer 53
604 CAAAT.....TGGCATCTGATGA..... 624
54 AspLysAlaGlnIleAspIlePheLysSerPheAspIleGlyGlnLysGln 70
625 .....GATGAGCCCAATACCGCGAAAGTTCATATCATATTCGAA 664
70 uValLysPheGlnGlnProAsnGlnLys.....AlaV 81
665 GTCCGATTTCTGGCTGTTGGCGCAATCCTTTCACAAATGCA... 711
81 alaIleLysAsnArgValIleGlyGlyAsnAlaSerGlnIleGlnGlyLys 97
712 ...TCAGTGTGGCAGACAGTCACTTACGTAGTGAATAATTAACATAG 758
98 LeuThrAlaAsnGlyLysValTyrIleVala.....As 108
759 CCCATATGTTTTCACACAGAGGCTCA..... 789
108 nProAsnGlyValIleIleThrGlnGlyAlaGlnIleAsnValAlaGlyL 125
790 .....TTGGCAGAGTGCCTCA 807
125 euleAlaThrThrLysAspLeuGlnuArgIleSerGluAsnSerAsnSer 141
808 CCAATGTTATCTATGATGCCCAAAAGCAAAAGTGAATTAATATGCG... 855
142 TyrGlnPheThrArgArgThrLysAspArgGlnValLeuLysGlnGlyLe 158
856 GTATATGCAACGGCG.....AACCCATATATGGAATAA 889
158 uValLeuLysAspGlyGlnValValLysGlnGlyGlnValIleAsnGln 175
890 GCAATGGCTCCAGCTGTTGCTGAAGATTGTTCTAT.....GAT 930
175 Lysn.....IleThrAlaGlnAspPheValValLeuAsnGlyLys 188
931 GAAATCTTGTGCTGAGATACCATTCAGTATTTACAGAACACAGCTCAAA 980

```

```
189 GlnValIleasnlysglyAsnIleasnValGluIlyAsnSerThrIleAs 11
188 GlnValIleasnlysglyAsnIleasnValGluIlyAsnSerThrIleAs 11
187 TGGGAATACTCTTTACGACCATATATGACACA..... 1017
186 NGIlyValIlyLeuSerSerIlyTyranPheThrPheThrLeuProA 222
1018 ..GGAATAATCAATGCCAAACATGAACACATCTCTGCTCAATGATTA 1065
222 spSerIlyIleSerValAlaLeuGluAspAsnThrValGlnIlyLeuAl 238
1066 AAACACAGA...ACGTCATATGTTTATGTTTCTTTATCCGACAGC 1112
239 LysAsnGlnIlySerIleIlySerIleGlyGlnIlyLeuSerAlaIlysgl 255
1113 AAGACACCTGTTTATCATGTGCA.....GGTGTGTCAACA 1150
255 YArgIlyGlnAlaLeuAspSerLeuValMetAsnAsnIlyValLeuGluA 272
1151 GTTATCGACCCGACTGAATAATGCA.....GAAAT 1182
272 LaThrIlyValSerAsnIlyAsnGlyIlyValIleuSerAlaAspAsn 288
1183 ATTTCCTTATTTGACGACAGA.....AAGGCGAATTGATA..... 1218
289 ValGlnIleuAsnAsnGlnSerAsnIleIlysglyIlyLeuValThrPheG1 305
1219 .....CTTACGACGAC.....ATCAATCAAGTGTGAGAGATTAT 1255
305 YAlaAspValThrSerAsnIlyGlnIleuIlyAspAsnIleIlyS1LeThrs 322
1256 ATTTCGACGAGATTTCGTCGCTCGCTGAATAATGAACGAACTGGCA 1305
322 eRlyThrGlySerIlyValThrSerProIlyIleAsnPheThr..... 336
1306 GGCAGCGGCTCATATCACT.....GAAGACAGTACCT 1340
337 GlyIlySerValAsnIleAsnGlnIlyAsnPheGlyArgIlyAspSerThrTh 353
1341 TACTGGAAAGTAACCGCTGCGCAACGACCGCTGTCGCAAAATCGGC. 1389
353 rHsIlyIly.....AspIleuPheIlyIlyLeuAsnT 364
1390 .....AAAGCAGCGTGCACCTTCACGCC 1413
364 hrcIlyValAsnIleAspValProAspAsnGlnIlyAsnIleArgIleAlaAsp 380
1414 AAAGGCGAAACCGACGCTCGATCAGCGCGGCGACGACGATCATTTT 1463
381 IleGlnAspAsnThrGlyThrGlyThrGlyThrGlyThr..... 394
1464 GGATCGACGACGACGATAAAGGCAAAACMACCTTGTAGTGAATCG 1513
395 .....SerSerPheIleGlnThr 401
1514 GC.....TTGTCAGCGGCGAGGGTACGTCGCACTGAATGCC 1551
401 lYAlaLeuSerSerLeuLeuAlaAsnAsnGlyIlyValAsnLeuIlysgly 417
1552 GATATCACTTCAACCCGACAAACTATTTTC...GGCTTCGCGGC.. 1596
418 AsnAsnValAsnIleSerGlyArgIleHisIleAspSerPheArgIlyse 434
1597 .....GAGCTTTGATTTAACGGGC 1618
434 rAspSerLeuLeuIlySerThrAsnIlyGlnIlyHisIleAspIleAsnAsn 451
1619 ATTGCGTTTCTTCCACCGTATTTCAAAATACGATGACGAGGGGATGATT 1668
451 lAspIleHis.....SerIlyArgIlyLeuPhe 460
1669 .....GTCAACCAATCAAGCAAAAGATCCACGTTACCAT 1706
461 PheIleThrSerLeuGlnAsnGlnIlyAspPheIlySerAsnIleThrI1 477
1707 TACAGCAGATAA.....GATATTGCTACACCGCATATACACA 1747
477 eThrAspSerIlyIleAsnLeuGlyAsnGlnIlyAlaMetGlyLeuGlyArgS 494
1748 GCTTGATGACAAATAAAGAAATTCCTAC...AACGTTGTTGGCGAG 1794
494 eValAspIlyIlyAsp.....TyraSpAsnArgTrp..... 504
1795 AAAGATACGACCAAAACGACCGG.....CG 1820
505 .....GlnIlySThrGlnIlySerIleArgIlyIlySerPheAspValIy 518
1821 GCTCAACCTGTTTTCAGACCGCGCGGACGAGACCGACCTGCTGCTT 1870
518 smetSerAsnValGlnPheAsnGlnValAspAsp.....ValIleLeuA 533
1871 CCGCGGGA.....ACAAATTTAAACGACACATCAGCAACAAC 1911
533 lAgIlyIlyPheGlnIlyValAsnLeuAsp..... 542
1912 GGCACACTGTTTTCAGCGGCGACCAACCGCACGCTACATCATTTT 1961
543 ..LysIleValAlaThrGlyGln..... 549
1962 AAACGACCATTTGTCGCAAAAGAGGCGATTCTCGCGGCAATGCTGT 2011
549 ..... 549
2012 GGCACACGACGTGATACCCGACATTTAAAGCGAAATTCGCAATT 2061
550 .....ThrAsnPheIlyIle 554
2062 AAAGCGGACAGCGCGGTGTTTCCCGCAATGTTGCCAAAGTGAAA...CG 2108
555 AspIlyIly.....ValSerArgAsnGlyArgIlyIlyIlyIlyIly 568
2109 CGATTGGCATTTGACCAATCAGCCACCAACA..... 2139
568 YAlaLeuAspLeuAspIlyArgThrIleuSerGlnLeuAsnGlnIlyA 585
2140 .....GTTTGGTGTGCGACCGCATCAACGACACA 2172
585 rArgArgArgTrpGlyIlyTrpIlyAspLeuGlnLeuAspMetAsnArgAla 601
2173 ATCTGACACGTTGCGACTGACGCGGTCTGCACAAATGTGTGCAAAAC 2222
602 TyrLeuTyrArgPheAspLeuPheAlaThrIlyAsnThrGlyArgSerTh 618
2223 CATTCACGAC..... 2232
618 rIleIlyAspThrGlnIleAsnIleSerAsnSerAsnIleAsnLeuIlySA 635
2233 .....GATAAGTATGCTTCAATTGACTACAG 2259
635 snGlyPheValHisLeuLeuAlaGlnIlyIlySerIleuAspAsnSerIlyS 651
2260 ACCGACATCAGCGGCAATGTCAT.....CTGCGCATCA 2294
652 IleAspIleThrPheAspIlyAspAsnSerIleAspThrIleuAlaGlnTh 668
2295 CGTCATTTAAATCTCACAGG...CTTGCACACTCAACGCAATCTTA 2341
668 rAsnArgLeuGlnIlyMetAsnGlnIlyValSerMetIleAsnSerHisIle 685
2342 GTCCAAATGCGGATACGCTTATACAGTGCAGCACAACGCAACCCAAAC 2391
685 YsIleValGlyAspGlnIlyGlnIlyIlySerProThrGlyThrIlyAla 701
2392 GGCACCTT.....ACGCTGTGGGCAATGCCACGCAACATTTA 2432
702 ThrMetPheLeuIleGlyIlyLeuIleGlyIlyGlyIlySerSerIlePheVa 718
```

2433 TCAAGC.....ACATTAAAGGCAACATCGCCTT 2464
 ID WAPA_BACSU STANDARD; PRT: 2334 AA.
 718 IlySerHisGlnGlyThrPheLysThrAspGlyAsnThrLysIleA 735
 2465 CGGCAATGCTTCA...TTTAATCTAAGGACACCGCGTACAAACGGC 2511
 735 IagLysLysLysLysLysLysLysLysLysLysLysLysLysLys 751
 2512 AGTCGAGCGCTTTCGGCAACGCTAAGCAACGTAAGCCATTCCGCACT 2561
 752GlyArgAlaAlaGlnGluVal.....LeuI 760
 2562 CAACGTAATGTCCTCCTAGCGGTAAGCGCATTCATTGTAAGCA 2611
 760 eAsnGlyAlaLeuGlySerAlaAspAsnAspAla...AsnIleAlaAsn 776
 2612 GCGCGTTTACGAGCAATCAGCGCGGCAAGATACGCAATTACACTTA 2661
 776 eAlaPheThrIleGlyAspSerAlaAsnThrLysThrIleGluAsn 792
 2662 AAAGACGGAATGAGACCTCGCTCAGACGACGAAATTAGGCAATTAA 2711
 793 AlaAspIleThrAlaLeuAlaProAsnGly.....GlyThrAlaTy 806
 2712 CCTGCAACGCGCACCATTACACTC.....AATTCGCGCTATC 2749
 806 rLeuSerSerLysAspValGluIleGluValLysProAsnSerAsnPe 822
 2750 GCCACGATCGGCGAGGCGCAACCGGACGACGACGAGTGGCGGCG 2799
 823ThrPhePheGluLeuProArg 829
 2800 CGCGGTTGCGCGCGTTCGCGCGCTTCCTATTATTCGTTACACGCCA 2849
 829 829
 2850 TTGCGTAGAATCCGTTTACACAGCTGACGGTAACGCC.....A 2890
 830GluLysAsnLeuAsnGlnThrLysIleAsnGlyLysLeu 844
 2891 AATTGACGCTCAGGGAACATTCGCTTATGTCGGAACCTTCGCGCTAC 2940
 844 yLysLeuSerGluArgGlyPheAlaArgLeuLysLysLysLysLys 860
 2941 CGCAGCGCAAAATGAAAGCTGGCGGAAAGTTCGGAAGGCACTTACAC 2990
 861 ArgAlaSerAsnLeu..... 865
 2991 GCGCGTCAACAATACCGCAACGCTGCAAGCTCGAACAATTCAGCG 3040
 866SerAlaGlnGlnLeuAsn 872
 3041 TAGTGAAGAAAGAAACACAAACCGCTGCCGAAACCTTATTCAC 3090
 872 alThrAspAlaSerGln...LysIleIleAsnThrLysLeuValSer 887
 3091 CTGCAAAACGAACACGTC.....GATCGCGCGCGCTG 3122
 888 LeuAspValGluLysLeuValSerValAlaValLysAspIleGly 902
 3123 GCGTTACCAACTCATCCGCAAAAGCGGAGTTCGCTGATTAATCGG 3172
 903AsnGlyC 905
 3173 TCAAGACAAAGAGCTTCGCGAANA.....CTC 3201
 905 yseGluGlnGlnGlnPheGlyAspLysGlyAsnAsnThrLysValSer 921
 3202 GCGAAGCAGAAAGCAAAAAA 3222
 922 GlyGluLeuGluAlaGluGln 928

seq_name: SwissProt_40:WAPA_BACSU
 seq_documentation_block:
 ID WAPA_BACSU STANDARD; PRT: 2334 AA.
 AC 007833;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Wall-associated protein precursor.
 GN WAPA OR N176.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168; Pubmed=8316082;
 RX MEDLINE=93302506; Pubmed=8316082;
 RA Foster S.J.;
 RT "Molecular analysis of three major wall-associated proteins of a Bacillus subtilis 168: evidence for processing of the product of a gene encoding a 258 kDa precursor two-domain ligand-binding protein."
 RT Mol. Microbiol. 8:299-310(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168 / BGSCL1;
 RX MEDLINE=95219088; Pubmed=7704263;
 RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
 RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis genome containing the hut and wapa loci."
 RT Microbiology 141:337-343(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168 / BGSCL1;
 RX MEDLINE=97124196; Pubmed=8969509;
 RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N., Miwa Y., Fujita Y.;
 RT "Sequencing of a 65 kb region of the Bacillus subtilis genome containing the hlc and cel loci, and creation of a 177 kb contig covering the gut-sacry region."
 RT Microbiology 142:3113-3123(1996).
 CC -1- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM, MOTILITY, SECRETION OR DIFFERENTIATION.
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED INTO THE MEDIUM.
 CC -1- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS; THE N-TERMINUS, HAS THREE 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED MOTIF REPEATED 31 TIMES.
 CC -1- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME SIMILARITY TO THE REPEAT IN E.COLI RBS GROUP OF PROTEINS (RHS-A-D).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC
 DR EMBL; L05634; AAA22883.1; -
 DR EMBL; D31856; BAA06656.1; -
 DR EMBL; D29985; BAA06260.1; -
 DR EMBL; D83026; BAA11683.1; -
 DR EMBL; Z99124; CAB15959.1; -
 DR PIR; S32920; S32920.
 DR Subtilisin; BGI0797; WAPA.
 DR InterPro; IPR003305; CBD_6.
 DR Pfam; PF02018; CBD_6; 1.
 DR Cell wall; Repeat; Signal; Complete proteome.
 KW SIGNAL 1
 FT OR 32 (POTENTIAL)
 CHAIN 29 2334 WALL-ASSOCIATED PROTEIN.

```

FT DOMAIN 504 869 3 X 101 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 504 605 1-1.
FT REPEAT 636 736 1-2.
FT REPEAT 769 869 1-3.
FT DOMAIN 1021 2139 31 X 21 AA APPROXIMATE TANDEM REPEATS OF
X(4)-G-X(4)-(YF)-X-D-X(2)-G-X(4).
FT REPEAT 1021 1040 2-1.
FT REPEAT 1042 1061 2-2.
FT REPEAT 1063 1082 2-3.
FT REPEAT 1083 1102 2-4.
FT REPEAT 1109 1128 2-5.
FT REPEAT 1129 1148 2-6.
FT REPEAT 1150 1169 2-7.
FT REPEAT 1174 1193 2-8.
FT REPEAT 1199 1218 2-9.
FT REPEAT 1219 1238 2-10.
FT REPEAT 1246 1265 2-11.
FT REPEAT 1667 1686 2-12.
FT REPEAT 1690 1709 2-13.
FT REPEAT 1711 1730 2-14.
FT REPEAT 1732 1751 2-15.
FT REPEAT 1753 1772 2-16.
FT REPEAT 1795 1814 2-17.
FT REPEAT 1820 1839 2-18.
FT REPEAT 1840 1859 2-19.
FT REPEAT 1861 1880 2-20.
FT REPEAT 1887 1906 2-21.
FT REPEAT 1908 1927 2-22.
FT REPEAT 1929 1948 2-23.
FT REPEAT 1969 1982 2-24 (APPROXIMATE).
FT REPEAT 1983 2002 2-25.
FT REPEAT 2008 2027 2-26.
FT REPEAT 2051 2070 2-27.
FT REPEAT 2071 2090 2-28.
FT REPEAT 2093 2112 2-29.
FT REPEAT 2120 2139 2-30.
FT REPEAT 2139 2158 2-31.
SO SEQUENCE 2334 AA: 258329 MW: B75138CCD278BA3 CRC64.

```

```

alignment_scores:
  Quality: 208.00      Length: 1236
  Ratio: 0.382        Gaps: 63
  Percent Similarity: 44.013    Percent Identity: 19.822

```

alignment_block:

US-09-303-518d-649 x WAPA_BACSU ..

Align seg 1/1 to: WAPA_BACSU from: 1 to: 2334

```

160 TATCCGACTTGGCGAAATATAAGCAAGTTTCCAGTGGCGGGAAGA 209
    ||| : : : : : : : : : : : : : : : : : : : : : : : :
988 TTTLeuGluLeuThrGluThrAlaSerPheLeuLeuLeuThrLys 1004
210 T.....: : : : : : : : : : : : : : : : : : : : :
1004 peltThrSnaLatTyrPheSnlLysGlyGlyLysLeuGlnLysVal 1021
211 .....ATTGAGTTTACACAAAGGAGGAG 237
1021 aLasrGlyHisAsnAlaThrValTyrThrGlyHisAsnLysGln 1037
238 TTGGTGGCAATCATGACAAAGCCCGATGATTTCTGTGCT 287
    ||| : : : : : : : : : : : : : : : : : : : : :
1038 LeuThrAlaLeuThrAspAlaSerGlyArgLysLeuThrPheThrLys 1054
288 GTCCGCTAAGCGCTGGCGCATTTGGGAGC.....: : : : : 318
    : : : : : : : : : : : : : : : : : : : : :
1054 pGluAsnGlyHisValThrSerLeuThrGlyPheLysAsnLysVal 1071
319 .....GATCAATATATTGAGCGCTGGACAT...AAGCGC 351
    : : : : : : : : : : : : : : : : : : : : :
1071 hrTyrSerTyrGluAsnAspLeuLeuLysValThrAspThrAspGly 1087

```

```

352 GGCTATTAACACGTTGATTTGGTGGAGAGGAAGAAATCCGATCAACA 401
    : : : : : : : : : : : : : : : : : : : : :
1088 ThrValThrSerTyrAspLysAspSerGluThrGlyLeuValLysGln 1104
402 TCGT.....TTTACT.....: : : : : 411
    : : : : : : : : : : : : : : : : : : : : :
1104 rSerLAsnSerThrGlnAlaLysProValPheThrGluThrGlnTyr 1121
412 .....TATTAATTTGTGAACGAAATATTATAAGCAGCGACTAAAGC 456
    : : : : : : : : : : : : : : : : : : : : :
1121 ergLynHisArgLeuGlnLysAlaLeuSnaLysLysGluThrVal 1137
457 CATCTTATGGCGCGAT.....TATCATATGGCGGCTTT 491
    : : : : : : : : : : : : : : : : : : : : :
1138 TyrSerTyrAspAlaSerLysLysThrLeuLeuMetThrGlnProSna 1154
492 GCATTA.....TTTG 502
    : : : : : : : : : : : : : : : : : : : : :
1154 YArgLysValGlnTyrGlyTyrAsnGlnAlaGlyAsnProIleGlnVal 1171
503 TCACAGATGCGAAGCCGTGTAATGACCGATTATGAGCGCGGA 552
    : : : : : : : : : : : : : : : : : : : : :
1171 LeAspAspAlaGluGlyLeuLysIleThr.....AsnThrLys 1184
553 TATATGATCAAAATTAATTAC.....CTGACCGGTGTGC 587
    ||| : : : : : : : : : : : : : : : : : :
1185 Tyr...GluGlyAsnAsnValValGluAspValAspProAsnAspVal 1200
588 TATTGGGCGAGCGCAATATGGCGATCTGATGAAGTGAAGCCAA 637
    ||| : : : : : : : : : : : : : : : : : :
1200 YThrGlyLysAlaThrGluSerTyrGlnTyrAspLysAsp.....Gly 1215
638 ACCGCCAAAGTTCATATGATGCAAGTGTCTGCTGCTGTGT 687
    ||| : : : : : : : : : : : : : : : : : :
1215 snValThrSer.....ValLysAspAlaTyr.....Gly 1224
688 GGCATATCCTTGCACAAATGATCAGGTGGTGCACAGTCATAGG 737
    : : : : : : : : : : : : : : : : : :
1225 ThrGluThrTyrGluTyrAsnLysAsnAspValThrLysMetLys 1241
738 TAGTGA.....AAATTAACATAGCCCATATGTTTATACACAGAG 784
    : : : : : : : : : : : : : : : : : :
1241 pThrGluGlyAsnValThrAspIleAlaTyrAspGlyLeuAspAlaVal 1258
785 GGTCAATTTGGCGACAGTGC.....TCACCAATGTTATCTATGAT 825
    : : : : : : : : : : : : : : : : : :
1258 ergLutThrAspLeuSerGlyLysSerSerAlaAlaValTyrAspLys 1274
825 .....: : : : : : : : : : : : : : : : : : 825
1275 TyrGlyAsnGlnIleGlnSerSerLysAspLeuSerAlaSerThrAsn 1291
826 .....GCCAAAGCAAAAGTGGTATTAAT 852
    ||| : : : : : : : : : : : : : : : : : :
1291 eLeuLysAspLysSerPheGlnAlaGlnLysSerGlyTyrAsnLeuThr 1308
853 .....GGGTATTGCAACGCGCAACCCCTAT 879
1308 lAserLysAspArgArgLysIleSerValIleAlaAspLysSerGlyVal 1324
880 ATAGAGAAAGCAATGCTTCACAGTGGTGTGTAAGATGCTTC..... 924
    : : : : : : : : : : : : : : : : : :
1325 LeuSerGlySerLysAlaLeuGluValLeuSerGlnSerThrSerAla 1341
925 .TATGATGAATCTTGGTGGAGATACCAATGATGATTTCAAGAAC 973
    ||| : : : : : : : : : : : : : : : : : :
1341 YThrAspHisGlyTyrSerSerAlaThrGlnThrValGluLeuGluPro 1357
974 GTCAAAATGGGAATACCTTTTAAGCAGCATATTAATGCGACAGAA 1023
    ||| : : : : : : : : : : : : : : : : : :
1358 .....AsnThrThrTyrThrLeu.....SerGlyLys 1366

```

```
1024 ATCAATCCCAACATGAAACAATCTCTGCTTAATAGATTAAACACG 1073
|||||:
1367 ILeuysThr.....AspleuAlaLysSerAr 1375
1074 AACGGTCATTTGTTAATGTTCTTATCCGACAGACGACGAAAGACCTG 1123
|||||:
1375 gAla.....TyrPheAsnIleAspleuArGAspLysAspGlnLysArgI 1390
1124 TT.....TATCAGCTCCAGGTGGTGTCAACATTTAT 1155
|||||:
1390 legIInPrIleIleAsnGluTyrSerAlaLeuAlaGluLysAsnAspTrp 1406
1156 CGACCCGAA.....CTGAATATGAGAAATAT 1184
|||||:
1407 ThrLysArGAlnIleThrPheThrProAlaAsnAlaGluLysAlaVal 1423
1185 TTCCCTTATTTGAC.....GAAGCAAAAGCGAATTGA 1216
|||||:
1423 lValTyrMetGluValAspHisLysAspLysAspGluLysGluLysAlaT 1440
1217 TACTTACCAGC...AACATCAATCAAGGTGCTGAGGATATATTTCCAA 1263
|||||:
1440 rPheAspGluValGlnLeuGluLysGly.....Glu 1450
1264 GGAGATTATACGCTCGCTGAAATTAACGAACTTGCGACGCGCGG 1313
|||||:
1451 ValSerSerSerTyrAsnProValGlnAsnSerPheThrSerAla.. 1466
1314 CGTTCAATACATGAAAGACAGTACGTTACTTGGAAATGAACGCGG 1363
|||||:
1467 .....ThrGluAsnTrpAsnValSerGluAlas 1476
1364 CA.....AACGACCGCGCTGCCAAATGCGCAA 1392
|||||:
1476 eValAspSerGluGluGlnPheAsnAspValSerLeuLysAlaIa 1492
1393 GGCAGCTGCACGTTCAAGCCAAAGGGGAAACCAAGCGTCGATAGCGT 1442
|||||:
1493 ArgThrSerAlaSerGlnAlaGly.....SerVa 1502
1443 GGCAGCGATGATCATTTTGGATAGAGCAGACGATTAAGCCAA 1492
|||||:
1502 lThrLysGlnThrValValLeuGluLysSerAlaAsnAspLys..... 1516
1493 AACAGCCTTTAGTGAATCGCTTGTCAGCGGACGCGGTACGTCGAA 1542
1516 ..... 1516
1543 CTGAATGCCGATATACATTCAACCCGACAACTATTTGGCTTCG 1592
|||||:
1517 .....ProValTyr..... 1519
1593 CGCGGACGTTTGATTTAAAGCGGCTTGGTTGCTCCACCGATATTC 1642
|||||:
1520 .....LeuThrLeuThrGlyMetSer..... 1526
1643 AAAATACGATGAAGGGCGCATGTGTCAACACCAATCAAGCAAGAA 1692
|||||:
1527 .....LysAla 1528
1693 TCCACCGTTACCATTAAGCAATTAAGATTTCTACACCGGCAATA 1742
|||||:
1529 SerSerValLysPheThrAspGluLysPyrSerLeuGlnAlaAsnVa 1545
1743 CAACAGCTTGATAGCAAAAGAAATGCTTACACGCTTGCTT... 1788
|||||:
1545 lThrTyrAlaAspGlySerThrGlyIle...TyrAsnAlaLysPhePro 1561
1789 ..GGCGAGAAAGAT.....ACAGCCAA 1809
|||||:
1561 eArglTyrGlnGluTrpAsnArgAlaAlaValAlleProLysThrLys 1577
1810 ACAGACGGGCGG.....CTCAACTTGTTCACAGCCGCGGAGAA 1853

1578 proIleAsnLysValAspIleSerIleLeuPheGlnLysSerAlaThr.. 1593
|||||:
1854 CCGCACCTTGCTGCTTTCGCGGCAACAATTTAAACGCAACATCACGC 1903
1594 .....GlyThrValTrpPheAspIleArgL 1603
1904 AACCAACGGCAACCTGTTTTCAGCGGCGACACCAACCGGACCTTAC 1953
|||||:
1603 euIleGluGlySerLeu.....ThrysserThrTyr 1614
1954 AATCATTTAAACGACCACTTGGTCGCAAAAGAGGCGATTCCTCGGGGA 2003
|||||:
1615 AspSer...AsnGluAsnTyrValTrpLysGluGluAspGluLeuGly... 1629
2004 AATGCTGTGGACACAGCAGTCGATCAACCGCACATTTAAAGCGAAGCT 2053
|||||:
1630 .TyrAlaThrSerThrAspTrpAspGluTrpGluLysLysThrSerGluT 1646
2054 TCCAAATTAAGCGGACAGCGGCTGCTTCCGCAATGTTGCCAAAGT 2103
|||||:
1646 hrAspAlaLysGluGluLysThrTyrThrTyrThrAspGlnAlaAspGln 1662
2104 AAAGCGATTGGCATTTGACCAATCAGCCCAAGCAAGTTTGTGCTGCG 2153
|||||:
1663 LeuThrAsnMetThrLeuSerAsn..... 1670
2154 ACCGATCAAAAGCCACCAATCTGTACAGCTTCGACCTGACGGGTGTA 2203
|||||:
1671 .....GlyThrSerIleLeuHisSerTyrAspLysGluGly... 1682
2204 CAATTTGTGTGCAAAACCAATTAC... 2229
|||||:
1683 ..AsnGluValSerLysThrIleArgAlaGluAlaAspGlnThrTyrLys 1698
2230 ...GACGATTAAGTGATGTCTCATTTGACTGACTAG...ACGCACTACGCG 2273
|||||:
1699 rPheGluTyrAspValMetGlyLysLeuValLysThrTrpAspProLeuGlu 1715
2274 CAATTTGATCTTCCGATCAC...GTCATTTAAATCTCACAGGCTTG 2320
|||||:
1715 yAsnValIleuAlaSerGluTyrAspAlaAsnSerIleuThrLysThrI 1732
2321 CCGCACTACAGCGCAAT.....CTTAGTGAATGCGCATACAGCT 2361
|||||:
1732 leSerProAsnGlnLysGluValSerLeuSerTyrAspLysThrAspArg 1748
2362 TAT...ACAGTCAAGCCACAGCCACCAA..... 2388
|||||:
1749 ValLysSerLysSerTyrAsnGlnThrGluLysTyrIlePheThrTyrAs 1765
2389 ...AACGGCAACCTTAGCTCTGGGCAATGCCCAAGCAACATTTAATC 2434
|||||:
1765 pLysAsnGlnLysGlnThrSerValValAsnLysGlu.....G 1778
2435 AAGCCACATTTAAACGCAACATTCGCTTGGGCAATGCTTCAATTAT 2484
|||||:
1778 lAsnThrThrLysLysArgThrPheAspAsnLysAsnArgLeuThrGlu 1794
2485 CTAAAGCACCGCGGTACAAAGCGCATCTGACG..... 2520
|||||:
1795 LeuThrAsp.....ArgGlyGlySerGlnThrTrpThrTyrProse 1808
2521 .....CTTCCGCGACAGCTTAAG 2539
1808 rAspSerAspLysLeuLysThrPheSerTrpIleAsnGlyAspGlnLysG 1825
2540 CA.....AACGTAAAGCATTCGCGCAACGATTAATGTCCTCCAGCC 2583
|||||:
1825 lYThrAsnGlnPheThrTyrAsnLysLeuAspGlnMetIleGluMetLys 1841
2584 GATAAGCA.....GTAATCCATTTGAAGACCGCGCTTACCGG 2624
|||||:

```

```

1842 AsperTherSertYrSerPheasPtyrAspGluAsnGlyAsnValG1 1838
2625 ACAATATCAGCGCGCAAGATACGGCATTTACACTTAAAGACAGCGAAT 2674
1858 nThrPheIleThrGlyAsnGlyValGlyThrSerPheSertYrAspGluA 1875
2675 GAGAGCTGGCGTACGAGCAGCAATTAAGCAATTTAACTTGACAAAGCC 2724
1875 rGAsnLeuValSerSertLeuHISileGlyAsp.....LysAsnGly 1888
2725 ACCATTATACACTCAATTCGGCTATGCGCAGCATGCGGAGGCGCAAC 2774
1889 GLyAspIleLeuThrGlySertYrGlyThrAspGlyAsnGly..... 1902
2775 CGGACGTGCGACAGATGCGCGCGCGCGCTTGGCGCGCTTGGCGCGCT 2824
1903 ..... 1905
2825 CCTATTATCGCTTACACCGCACTTGGTGAATCCGCTTCAACACG 2874
1905 hrThrIleAsnSerSertYrAspGlyValGlyThrGlyTyr..... 1919
2875 CTGACGGTAAACGGCAATTGAC.....GCTCA 2903
1920 .....GlyTyrLeuAsnGlyLeuValGlySglThrHISGluAs 1932
2904 GGGACAA...TTCCGCTTTATGTCGGAACCTTTCGGCTACCGGCGGAC 2950
1932 pGlyThrValIleGlyThrGlyThrAspGlyPheGlyAsnArgGlyThrV 1949
2951 AATTGAAGCTGGCGGAAGTTCCGAAGGCATTTACACCTTGGCGGTCAC 3000
1949 aThrThrIleGlyAspGlySertYrThrValAsnIleSertPheAsn 1965
3001 AATACCGGCAACGACACTGCAACCTTGAACATTTACAGGTGAGGAAG 3050
1966 .....IleMetAsnGlyLeuThrValAsn.. 1974
3051 AAAAGACAAACCGCTGCGCAAAACCTTAATTTACACCTTGGCAAAAG 3100
1975 .....AspGlySertYrSerPheSertYrAspGlyAsnGlyA 1985
3101 AACAGCTGATCGCGCGCGCTTACCAACTCATCGCAAGAGCGG 3150
1985 smArgThrSertYrAspGlyPheThrThrThrPheAspGlyAsn 2001
3151 GAGTTCGCGCTGATATCGGTCGAAGACAGAGCTTTCGCAACAT 3200
2002 .LeuThrIleValThrIleGlySglThrAspGlyProPheIleThrTyrL 2018
3201 CGGCAAGGCAAGCAAGCAAGCAAGCGCGCAAGCAAGCGCAAGCGC 3250
2018 ySrtYrAspGlyAsnGlyAsnArgGlyIleGlyThrValAsnGlyLys 2033
3251 TTGACG 3256
2034 ValThr 2035

```

seq_name: SwissProt_40:PTS_SERMA

seq_documentation_block:

```

ID PTS_SERMA STANDARD; PRT: 1045 AA.
AC P09489;
DT 01-MAR-1989 (rel. 10, Created)
DT 01-MAR-1989 (rel. 10, Last sequence update)
DT 01-NOV-1995 (rel. 32, Last annotation update)
DE Extracellular serine protease precursor (EC 3.4.21.-).
OS Serratia marcescens.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Serratia.
OX NCBI_TaxID=615;
RN [1]
SEQUENCE FROM N.A., AND SEQUENCE OF 28-35 AND 407-408.

```

```

RC STRAIN-IFO 3046;
RX MEDLINE=06223815; PubMed=3011754;
RA Taniguchi N., Dozumi T., Beppu T.;
RT "Specific excretion of Serratia marcescens protease through the outer
RT membrane of Escherichia coli.";
RL J. Bacteriol. 166:937-944(1986).
RN [2]
RP PARTIAL SEQUENCE, AND PROCESSING.
RX MEDLINE=92348352; PubMed=1639760;
RA Shikata S., Shimada K., Kataoka H., Hornouchi S., Beppu T.;
RT "Detection of large COOH-terminal domains processed from the
RT precursor of Serratia marcescens serine protease in the outer
RT membrane of Escherichia coli.";
RL J. Biochem. 111:627-632(1992).
CC -1 SUBCELLULAR LOCATION: Secreted.
CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL, M13469; AAA26572.1;
CC PIR, A29840; A29840.
CC HSSP: O99405; IMPT.
CC
CC MEROPS: S08.094;
CC InterPro: IPR000209; Peptidase_S8.
CC Pfam: PF00082; Peptidase_S8; 1.
CC PRINTS: PR00723; SUBTILISIN.
CC PROSITE: PS00136; SUBTILASE_ASP; FALSE_NEG.
CC PROSITE: PS00137; SUBTILASE_HIS; 1.
CC PROSITE: PS00138; SUBTILASE_SER; 1.
CC KW Hydroxylase; Serine protease; zymogen; signal.
CC FT SIGNAL 1 27
CC FT CHAIN 28 645 EXTRACELLULAR SERINE PROTEASE.
CC FT PROPEP 646 1045
CC FT ACT_SITE 76 76 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 112 112 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 341 341 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC SO SEQUENCE 1045 AA; 112345 MW; 4924E50E4FF179C CRC64;

```

alignment_scores:

quality: 207.50 Length: 1237
Ratio: 0.382 Gaps: 65
Percent Similarity: 43.897 Percent Identity: 19.563

alignment_block:

US-09-303-518d-649 x PTS_SERMA ..

Align seg 1/1 to: PTS_SERMA from: 1 to: 1045

```

1162 AGACTGATATATGCA.....GAAATAT 1184
|||||:|||||
91 ArgLeuAspAsnGlySertYrAsnPheSertYrAspGlyAsnAspAsn 107
1185 TTCCCTTTATGACGAAGAAAGCGAATTGATCTACACAGCAATCA 1234
|||||:|||||
107 tSerPheGlyAspHisGlyThrHISValAlaGlyIleAlaIleAlaLysA 124
1235 ATCAAGGTGCTGA.....GGATTATATTCCAGGAGCTTTTACGGTC 1278
|||||:|||||
124 rGAspGlyAlaGlyIleHisGlyValAlaPheAspAlaAspIleIleGly 140
1279 TCGCCTGAAT..... 1290
|||||
141 ThrTyrLeuAsnAspTyrGlyAsnArgAsnGlyArgGlyGluLeuIleG1 157
1291 .....AACGAACCTTGGCAAGGCGCGCGCTTCATA 1321

```


2915 GCTTATGTCGGAACCTCTGGGCTACCGCAGACAAATGGAAGTGGCG 2964
620 TQPHLEVALVSLAIALEUAASPTLYRAGLYASNAIPVALGILEUSLY 636
2965 GAAAGTTCGGAAGGCACTTACACCTTGGGGGTCACCAATACCGGCAAGA 3014
637 ILEthraThraspAlaserPheaspSerleuAlaserThrgluAangl 653
3015 AACTGCA.....AGCTGCAACATTTGACGGTACTGGAAGGAAG 3055
653 ULyAlaValAlaIArgAlaValAlglUthrleuAsnAlaThrglu..... 667
3056 ACAACAACCGCTGTCCGAAACCTTAATTTACCTCGCAAAACGAACAC 3105
668ProValThrglu.....ThralAlaLysArgSerVal 677
3106 GTTCGATCCGGGGCGGGCGGTTACCACTATCCGAAGAAGCGGAGATT 3155
678 ALAlaIerAlaAlaIglUglUAlaAsnleuGlnSerAspIly..... 692
3156 CCGCTGCATTAATCCGGTCAAAACAAGACTTTCGACAAACCTGGCA 3205
693
3206 AGCGAGAACCCAAAAAACAGCGGGAACAAACAGCGGCAACGCTTGAC 3255
694 IuAlaIglAlaValAsnU.....Glu 701
3256 GCGCTGATTTGGCGCGCGGCGGATCCGTCGA..... 3288
702 ALaserIleValAlaIglYHisProIleTyrgIuSerPheleuGlyPheTh 718
3289..AAGACAGAAACGTTGCCGACCGCGCGCGGCGGAGCGGAGAA..A 3334
718 rSerAlaArgIuSerGlnGlnAlaThraIrgIuSerGlyGlnIleH 735
3335 ATGTGCGATTTGTCAGCGCGAGAGAAAGAAAAACGGGTGAGCGGAT 3384
735 ISAlaAspMetAlaSerAlaIglInIleAsnGluSerArgTyrgIu..... 749
3385 AAAGACACCGCTTGGCGAAMACGCGGAGGGAAGGGAACCGCGCGCTAC 3434
750 ArgAspThrAlaThrgIuArgIuSerGlnAlaIglU..... 761
3435 CACGCGCTTCCCGCGCGCGCGCGCGCGCGGCGGATTTGCCGACATGC 3484
762GlyArgArgThrAlaThraspIleLys..... 770
3485 AACCCCAACCGCAGCCCAACCGCAGCGCGATGATCAGCGCTTTGCGCGT 3534
770 770
3535 AATAGCGGTTTGAGTGAATTTTCGCGCAACGCTCAACAGCTTTGCGCGT 3584
770 770
3585 ACAGACGAATTAAGACCGCGATTTTCCGAGAACCGCGCAACGCGCTT 3634
771AlaAspAsp.....AsnGlyAlaIat 777
3635 GGACAAAGCGCATCCGCGACACCAAAACACTAGCGTTCCGAAGT...TTC 3681
777 TrpAlaLysIleuLeuGlySerThrIrgIuAlaSerGlyAsnAspAsnAla 793
3682 CGCGCTACCGGCAAAACAGACCTGCGCAAAATGGATGACAGAAAAA 3731
794 ThrgIyTrgIuThraserThrTyrgIuValLeuLeuGlyIleuAspSerG 810
3732 CCTC...GCAGCGGCGCGCTGCGCATCTGTTTCCGCAACCGCGACCG 3778
810 UleuPheGlyAspIyArgIuLeuGlyMetethrIrgIyTrhArgThr 826

3779 AAACACCTTTCGAGCAGCGCATTCGGCAACTGCGACGCTTGCCCAACGCG 3828
827SerleuAspIyGlyTyrgIuSerAspAlaHis..... 837
3829 GCGGTTTTCGGCAATACGGCATCGACAGGTTTACATCGGCATGACGCG 3878
838SerAspAsnTyHisIleuGlyLeuTyrgI 847
3879 GCGCGCGGATTTTTCAGACGCGCAACCTTTTCAGACGCGATCGAGCAAA 3928
847 YAspIyAspArgPhe.....GlyAlaIleuAlaIeuArgAlaIglIyThr 862
3929 TC.....CGCGCGCGGCTGCTGCATTCATTCAGCGCATTCAG 3960
862 YTTThrPrlHisArgIleAspPrlHisSerArgSerValAsnTyrgIAlaGln 878
3961GCAGATTCACCGCGCGGTTTTCGCGGATTCGCGCAT 3995
879 SerAspIrgIuLysAlaIyStrIyAsnAlaIaIghThrgIyGlnIleuPheI 895
3996 CGAACCACACATCGCGCGCAACCGCTATTGTGTCGCAAAAGCGGATTAC 4045
895 GGU.....SerGlyTyraAspIrgI 902
4046 GCTACGAAACGTCAATATCGCCACC...CCGCGCTTGATTCACACCG 4092
902 HTrSerAspAlaValaAsnleuGluProPheAlaAsnleuAlaTyThHis 918
4093 TACCGC..... 4098
919 TyThrArgAsnGluGluIleAsnGlnGlnGlyAlaIleAlaIleuArgI 935
4099GCGCGCATTAAGGCGAG 4114
935 YAspIyGlnSerGlnSerAlaThraIlaSerThrleuGlyLeuArgAla 952
4115 ATTATTCATTCAAACCGCGC..... 4134
952 SpThrgIuIrgIuThraspSerValaIleAlaIleuArgIyIleu 968
4135CAACAC.....ATTCCATCACGCTTATTTAGAGCT 4166
969 GlyTrgIuGlnHisGlnTyrgIyLysIleuGluArgIyThrgIuIleuMetPh 985
4167 GTTCATACCGATCGCGCTCGGCAAAAGTCGAACACGCGTCATACCG 4216
985 ElyAspIrgThraspAlaIaIaIeAsp.....ValAsnSerV 997
4217 CCGATTCGCTCAGGATTTTCGCAAAACCGCAATGGGAAATGGCGGTA 4266
997 AlProValSerArgAsp...GlyAlaIleuIyAla.....GlyVal 1010
4267 AACCGCAATC...AAGGTTTTCAGGTCGCTCCGCAACGCTGCGCGCGC 4313
1011 AspValSerIleAsnLysAsnAlaValaIleuSerIleu.....G 1023
4314 CAAGCGCCCACTGGAAGGCCAACAC...AGCGCGGCAATCAAAATTAG 4360
1023 YTyrgIyGlyGlnIleuSerSerAsnHisGlnAspAsnSerValaIaIag 1040
4361 GCTACCGCTGG 4371
1040 IyLeuThrThr 1043

seq_name: SwissProt_40:YGY3_HAISO
seq_documentation_block:
ID YGY3_HAISO STANDARD: PRT; 437 AA.
AC P21561.
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-OCT-2021

385 SATG.....GlyArgLeuArgValArgGly 394

4057 GTCAATATCGCACCCCGGCTTGATTCACCGCTACCGCGCGCAT 4106

394 erThralaValProArg.....ProLeuProArgGlnHis 406

seq_name: Swissprot_40:VG37_BPT2

seq_documentation_block:

ID VG37_BPT2 STANDARD: PRT: 1341 AA.

AC P07067;

DT 01-APR-1988 (Rel. 07, Created)

DT 01-APR-1988 (Rel. 07, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Long tail fiber protein p37 (Protein Gp37) (Receptor recognizing

protein).

GN 37.

OS Bacteriophage T2.

OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;

OC T4-like phages

OX NCBI_TaxID=10664;

RP [1]

RA MEDLINE=87112716; PubMed=3806672;

RA Riede I., Drexler K., Eschbach M.-U., Henning U.;

"DNA sequence of the tail fiber genes 37, encoding the receptor

recognizing part of the fiber, of bacteriophages T2 and K3.";

J. Mol. Biol. 191:255-266(1986).

CC -1- FUNCTION: STRUCTURAL COMPONENT OF THE DISTAL-HALF TAIL FIBER.

CC -1- SUBUNIT: THE DISTAL HALF-FIBER CONTAINS TWO MOLECULES EACH OF GP36

CC AND GP37 AND ONE MOLECULE OF GP35.

CC -1- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC EMBL; X04442; CAA28038.1; -

DR PIR; S09579; S09579.

KW Fiber protein.

SO SEQUENCE 1341 AA; 143152 MW; 187981E7E5FA7F5 CRC64;

alignment_scores:

Quality: 205.50 Length: 1576

Ratio: 0.317 Gaps: 83

Percent Similarity: 41.117 Percent Identity: 18.718

alignment_block:

US-09-303-518D-649 x VG37_BPT2

Align seg 1/1 to: VG37_BPT2 from: 1 to: 1341

181 AAAGCAAGTTGACAGTCGGGCGAAGATATTAGGTTTACAACAAA.. 228

26 GIUGLIGLULeuAlaIleAsnLeuLysAspArgThrIlePheThrLysAs 42

229AAAGGAGAGTTGTCGGCAATATGACAAAGCCCGCATGTTG 274

42 PASPserGlyAsnIleIleAspLeuSerIleSerAlaGlyGlyAsnIle 59

275 ATTTTCTGTGTG.....TCGCGTAAGCGCGCGCGCA 309

59 erGlyAsnIleThrGlnThrGlyAspTyrThrGlnThrGlyLysPheAsn 75

310 TTGGTGGCGCATCATATATTGTCGGTGGACATATACGCGCGCATATA 359

76 LeuIleGlyProGlnIleValAlaSer.....GlyGlyTyr... 87

360 CAACGTTGATTTGGTGGGAGAGAGAAATCCGATCAACATCGTTTAA 409

88IleGluPhe..... 90

410 CTTATATAATTTGAAAGCAATATTATTAAGCGAGCATTAAGCCAT 459

91AsnTyrArg.....ThrThrGlySer 97

460 CCTATGCGCGCGGATTATCATATGCCGTTTGATTAATTTGTC..... 504

98 GlyAlaIlePheSerGlyGlnHisThrAlaLysAlaProIlePheValAsp 114

505ACAGATCGCAACCTGTTGAATGACCAAGTT 535

114 userSerAlaIleThrSerThrSerGluTyrAsnProIleIleLysGlnArg 131

536 ATATGATGCGCGG..... 549

131 helYAspGlyThrPheSerLeuGlyThrLeuValSerGluGlySerLeu 147

550AATATATGATCAATTAATTAATTAATTAATTAATTAATTA 593

148 LysIleHisTyrIleAsnGluSer..... 155

594 GCGACGCGAGCAATATTCGCGATCTGATGAAGATGACCCCAATACCCGC 643

156 .GlyAspSerLysTyrTP..... 161

644 AAAGTTATCATATTCGCAAGTCGATTCGTCGTCGTCGTCGTCGTCG 693

161

694 ACCTTTCACAAAATGATGATGATGATGATGATGATGATGATGATGAT 743

162 ThrIleHisTyrIleAsnGluSer.....GlyGlyPheThrValAspGlyGly..... 174

744 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 793

175GlyLeuGlyValSerGlyGlySerIleT 184

794 GCGACACTGTCG.....TCACCAATGTT 816

184 hrThrSerGlyAsnIleAlaIleAlaLeuGlyAsnIleThrSerProGlnIle 200

817

201 AsnThrLysAsnIleIleLeuAspThrLysAlaPheGlyGlnTyrAspSe 217

828 CCAAAAGCAAAAGTGTAAATTAATGAGGTATTCGAAAGCGCAACCCCT 877

217 rGln.....SerLeuValGlnTyrValTyrProG 227

878 ATATAGGAAAGCAATGCTTCACGCTGTCGTAA..... 915

227 LyrThrGlyGluLysnGlyIleAsnTyrLeuArgLysValArgAlaLys 243

916GATTGGTTCATGATGAATCTTCGTCGACATACCCCTTACGT 959

244 SerGlyGlyThrIleTyrHisGlnIleAlaSerAlaGlnThr..... 257

960 ATTTCAGCAACACAGCTCAAAATGAGAAATCTTTTACGACAGATAATA 1009

258GlyLysAsnAspGlnIleSerTyrPheThrGlyAsnT 270

1010 ATGGCACAGAAATAATCAATGCGCAATAGACACATTCCTGCTTAAT 1059

270 hrLeuThrThrLysLeuMetCylLeuArgAsnAspGlyAlaMetValLeu 286

1060 AGATTAAACACGACACCTGATTAATGTTTATTTCTTTATTCGACAGAC 1109

287 Arg.....ArgSerLeuAlaIleGlyThrIleThrAlaAspLeu... 299

1110 AGCAAGAGAACCTGTTTATCATCTGTCAGGTGTCGACAGATTATCGAC 1159

[illegible][illegible]

[illegible]

3542 GGCACCCCAACCCACGCCAACCGAGCGGACCTGATACGCCGTTT 3551
 1025 euArgPro.....Phe 1028
 3532 GCC.....AATAGCGGTTTAGTGA 3551
 1029 AlalleSnmetalaaSpglYatgValglYmetlaSnhlglYleuAaSnll 1043
 3552 ATTTTCGCCACGCTCAAC.....ACCGTTTCGGCGTAC 3586
 1045 ethGlYglYglYleuAaSnValThrGlYglYleuAaSnValThrSerGlY 1062
 3587 AGACGAAATTAGAACCGGATTTTGGCGAAGACCGCGCAACGCCGTTTG 3636
 1062 snThrSerleuGlYaaSnlle.....SerSerArgValValAlaArgTrp 1076
 3637ACAAGCGGCATCCGGGACACCAACACACTACCGTTGGCAAGATT 3680
 1077 ArgGlYAlaSerlYlTrpAlaSpaSnSerApThrMetLysSerLysll 1093
 3681 CCGCGCCCTACCGGCAACACCGACCTGCCAA..... 3714
 1093 ethrPheMetAlaSpaSnlglYaaSpLeuSerAaSerApSerlYlTrp 1110
 3715ATCGGTATGCGAGAAAACCTCGCGACGCGCGGTGGCATCTG 3759
 1110 rolleValglYAlaTrYserAaSnTrYglYserAlaGlYTr..... 1123
 3760 TTTTCGCAACCGGACCGGAAACACTTCGACACGCGATCGGCACTC 3809
 1124ArglnThrPheGlInPheLly..... 1130
 3810 GGCAGCGCTGCCCAACGCGCGCTTTTCGGGCATACGCGATCGACAGT 3859
 1130 1130
 3860 TCTACATCGCATCAGCGCGCGCGGCTTTTACGACGCGACGCTT... 3906
 1131 ..TrpValGlYSerGlYThrThrAlaGlYlTrpArgAaSpglYllelleArg 1146
 3907TCAGACGCGCATCGAGGCAAAATCCGCGCGCGTGGTCATTA 3950
 1147 lleArglYaaSpaSnAlaAaSnGlIn..... 1155
 3951 CGGCATTCAGGCGACGATACCGC..... 3972
 1156GlInAlaArgTrpArgPheThrMetAaSpglYThrleuAaCysP 1170
 3972 3972
 1170 roglYlYsValleuLeuProGlInThrGlYAlaPheGlYAlaSnThrSer 1186
 3973 GCCGTTTCGGCGGA.....TTTCGC..... 3993
 1187 AaSnGlYleuAaSnlYlYaaSnSerlleThrPheGlYaaSerApThrGlYll 1203
 3994ATCGAACCGGACATTCGGCGCAACGCGCT 4021
 1203 elYsGlInAaSnGlYaaSpglYleuLeuAaSpLlelYrAlaAaSnSerValGlInV 1220
 4022 ATTTC...GTCCAAAAGCGGAT...TTCGCTACGAAACGTCAATATC 4065
 1220 alPheArgPheGlInAaSnGlYaaSpLeuTrYSerTrYlYsAaSnlleAaSn... 1235
 4066 GCCACCCCGCGCTTCGATTCACGCGCTAC...CGCGCGGCGATTAAAGC 4112
 1236AlaProAaSnValTrYlleArgSerAaSpLleArgle 1247
 4113 AGATTATCATTCAAACGCGGCAACG 4140
 1247 ulYsSerAaSnPheYsProlleGlInAaSn 1256

```

seq_documentation_block:
ID P3P_LACLC STANDARD: PRT: 1902 AA.
AC P13292;
DT 01-APR-1990 (rel. 14, created)
DT 01-APR-1990 (rel. 14, last sequence update)
DT 16-OCT-2001 (rel. 40, last annotation update)
DE P11-type proteinase precursor (EC 3.4.21.36) (Lactocarpin) (Cell wall-
associated serine proteinase).
DE associated serine proteinase).
GN P11P.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.
RC STRAIN=SK1;
RA MEDLINE=89340435; PubMed=2760036;
RX Vos P., Simons G., Stezen R.J., de Vos W.M.;
RT "Primary structure and organization of the gene for a procaryotic,
cell envelope-located serine proteinase.";
RL J. Biol. Chem. 264:13579-13585(1989).
CC -1- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
GROWTH OF THE BACTERIA ON MILK.
CC -1- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
specificity, although some substrate preference have been noted,
e.g. large hydrophobic residues in the p1 and p4 positions, and
pro in the p2 position. Best known for its action on caseins,
although it has been shown to hydrolyse hemoglobin and oxidized
insulin B-chain.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
SUBTILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J04962; AAA03533.1; ALT_SEQ.
DR PIR: A32634; A32634.
DR HSSP: P00782; 2SPT.
DR MEROPS: S08.019; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR003137; PA.
DR InterPro: IPR000209; Peptidase-S8.
DR Pfam: PF00746; Gram_pos_anchor. 1.
DR Pfam: PF02225; PA. 1.
DR Pfam: PF00082; Peptidase_S8. 3.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP. 1.
DR PROSITE: PS00137; SUBTILASE_HIS. 1.
DR PROSITE: PS00138; SUBTILASE_SER. 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING. 1.
DR Hydrolase: Serine protease; Cell wall; Zymogen; Signal; Plasmid;
KW Transmembrane. 1
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1902
FT DOMAIN 188 1876
FT TRANSHEM 1877 1895
FT DOMAIN 1896 1902
FT ACT_SITE 217 217
FT ACT_SITE 281 281
FT ACT_SITE 620 620
FT DOMAIN 1867 1872
FT SEQUENCE 1902 AA: 200550 MW: 870CBAA9345F9D3 CRC64:

```

```

alignment_scores:
Quality: 205.00 Length: 1685
Ratio: 0.285 Gaps: 85
Percent Similarity: 42.730 Percent Identity: 19.525

alignment_block:
US-09-303-518D-649 x P3P_LACLC
Align seg 1/1 to: P3P_LACLC from: 1 to: 1902

124 GCGGACACACTTATTTGCGATCATCATCACTATCTATCGGACTTTC 173
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
387 SerGlySerAlaThrGluGlyValAlaSerGlyAspTyrTyr..... 399
174 CGAAATTAAGGCAAGTTTGCATCGCGGCAAGAAATATGAGTTTACA 223
400 .....GlyLeuGlnAspAsnGlnMetValG 408

224 ACAAAAAGGAGAGTGTGCGC.....AAATCATGACAAA 261
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
408 lserProGlyThrSerArgGlyAlaThrThrValAlaSerAlaGluAsn 424
262 GCCCGATGATGATTTTTCGTGTGTCGCGATACGCGCGCGCAT 311
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
425 ThrAspValIleThrGlnAlaValThrIleThrAspGlyThrGlyGlu 441
312 GGTGGCGCATATATATTTGTAGCGTGGCAGATTAAC.....GCGGCT 355
441 nleuGlyProGluThrIleGlnLeuSerSerHisAspPheThrGlySer 458
356 ATATAC.....AACGTGAT 369
458 heaspGlnLysLysPheTyrIleValLysAspAlaSerGlyAsnLeuSer 474
370 TTGTGCGGAGCAAGAAATCCGATCAACATCGTTTACTTAAAT 419
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
475 LysGlyAlaLeuAlaAspThrThrAlaAspAlaGlyLysIleAlaI 491
420 TGTGAAGGATATATTTATTAAGCAGGAGCTAAAGCCATCTTATGCG 469
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
491 eValLysArgGlyGluPheSerPheAspLys...GlnLysTyrAlaG 507
470 GCGATTATCATATGCGCGTTCATTAATTTGTCAGATGACAGTACCT 519
507 lAlaAlaGlyAlaAlaGlyLeuIleIleValAlaThrAspGlyThrAla 523
520 GTTGAATGACAGCTAT..... 537
524 ThrProMetThrSerIleAlaLeuThrThrThrPheProThrPheGlyLe 540
538 .....ATGGATGGCGGCAATATATATCAT.....CAAAATATATACCTG 577
540 userSerValThrGlyLysLeuValAspThrValThrAlaHisProA 557
578 AC.....CGTGTGCTATTTGGCGCAGCAGGCAATATGCGCATCT 618
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
557 spaSerLeuGlyValLysIleThrLeuAlaMetLeuProAsnGlnLys 573
619 GATGAGATGAGCCCATTAACCCGAAAGTTCATATCATATTCAGATGCG 668
574 TyrThrGlnAspLysMetSerAspPheThrSerTyrGlyProValSer 590
669 GTATCTTGG.....CTGTTGGTGGCAATACCTTGGCAGC 703
590 nleuSerThrLysProAspIleThrAlaProGlyGlyAlaIleThrSer 607
704 AAATGATCAGTGTGTCGACAGCACTAAGTATGAAAAATTTAA 753
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
607 hGlnAspAsnAsnGlyTyrThrAsnMetSerGlyThrSerMetAla... 622
754 CATGACCCATATGCTTTTACCAACAGAGGCTCA..... 769
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

623 ...SerProhe.....IleAlaGlySerGlnAlaLeuLeuLeu 634
790TTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAA 832
634 sGlnAlaLeuAsnAsnLysAsnProPheThrAlaTyrTyrLysGlnL 651
833 AGCAAAAGTGGTTAATTAAATGGGATATGCAAAACGGCAACCCCAATTA 882
651 eulysGlyThrAlaLeuThrAspPheLeuLysThr..... 662
883 GGAAGAAGCAATGGCTCCAGCTGGTCTGTAAGATTGGTCTATGATGA 932
663 ValGluMetAsnThrAlaGlnProLeu...AsnAspIleAsnTyr.... 676
933 AATCTTGGCTGGATACCATTCAGTATCTACGAACACCGTCAAAATG 982
677AsnAsnValIleValSerProTyrArgGlnG 687
983 GG.....AAATACCTCTTTAAAGACGATTAATATGAC 1014
687 lYAlaGlyLeuValAspValLysAlaIleAspAlaLeuGluLysAsn 703
1015 ACAGAAAAATCAATGCCAATGACATGAACAATCTCTGCTATAGATT 1064
704 ProSerThrValValAlaGluAsnGlyTyrProAlaValGluLeuLysAs 720
1065 AAAAAGCAGAACCGTTCATTTGTTTAAATGTTCTTTA..... 1101
720 PheThrSerThrAspLysThrPheLysLeuThrPheThrAsnArgThr 737
1102TCGAGACAGCAGACGACGACTGTT 1125
737 hrHISGluLeuThrTyrGluMetAspSerAsnThrAspThrAsnAlaVal 753
1126 TATATCTGCTGAGTGGTGTCAACAGT.....TATGACCCGAGCT 1166
754 TyrThrSerAlaThrAspProAsnSerGlyValLeuTyrAspLysLys 770
1167 GAATTAAT.....GGAGAAATATTTCTCTTATTTGACGAAG 1201
770 eaSpGlyAlaAlaIleLysAlaGlySerAsnIleThr...ValProAlaG 786
1202 GAAAGAGC.....GAATTCATCTTACCGACGACCAATCAAT 1236
786 lYThrThrAlaGlnIleGluPheThrLeuSerLeuProLysSerPheAsp 802
1237 CAAGGT.....GCTGAGATTTATTTCCAAAGAGATTTCAGCT 1277
803 GlnGlnGlnPheValGlnLysPheLeuAsnPheLysGlySer..... 816
1278 CTCGCTGAATAATTAACGAATTTGGCAAGCGCGGCGTCAATACACT 1326
817AspGlySerArgLeuAsnLeuProT 825
1327GAAGACAGTACCGTTACTTGGAAAGTA 1353
825 yMetGlyPhePheGlyAspThrAsnAspGlyLysIleValAspSerLeu 841
1354 AAGCGCGTGGCAAAACGACCGCTGTCNAATCGCAAA.....GGCAC 1397
842 AsnGlyIleThr.....TyrSerProAlaGlyLysAsnPheGlyThr 855
1398 GCTGACGTTCAAGCAAGGGAAGAAACGAGCTG..... 1434
855 rValProLeu...LeuLysAsnLysAsnThrGlyThrGlnTyrTyrGly 871
1435ATCAGCGTGGCGACGCTACAGTCAATTTGATCAGACGCA... 1476
871 lYMetValThrAspAlaAspLysAsnLysThrValAspAspGlnAlaIle 887
1477GACGATAAAGC..... 1488
888 AlaPheSerSerAspLysAsnAlaLeuTyrAsnAspIleSerMetLysTyr 904

1489AAAAACAAGCCTTTAGTGAATCGGCTGTCACGCGCA 1528
904 rTyrIleLeuAlaArgAsnIleSerAsnValGlnValAspIleLeuAspGly 921
1529 GGGGTACGGTGCACACTGATGCGGATTAATGAGTTCAACCCGACAACTC 1578
921 lncLysAsnLysValThrThrLeuSerSerSerThrAsnArgLysLysThr 937
1579 TATTCGGCTTTCGGCGGAGCGTTTGAATTTAAACGGCATTCGCTTTC 1628
938 TyrTyr.....AsnAlaHisSerGlnG 945
1629 GTTC..... 1632
945 nTyrIleTyrTyrAsnAlaProAlaTyrAspGlyThrTyrAspGlnA 962
1633CACCGTATTCNAATACCGATGAGGCGGATGATT..... 1668
962 rGAspGlyAsnIleLysThrAlaAspAspGlySerTyrThrTyrArgIle 978
1669GTCAACCAATCAAGACAAAGATCCACGCTTACATTACAG 1712
979 SerGlyValProGluGlyGlyAspLysArgGlnValPheAspAlaProph 995
1713 CAATTAAGATATT...GCTACAAACCGGCAATTAACACAGCTGGATAGA 1759
995 eLysLeuAspSerLysAlaProThrValArgHisValAlaLeuSerAla 1802
1760 AAAAAGAATTTGCTACACGCTTGGTT..... 1788
1012 yStrGluAsnGlyLysThrGlnTyrTyrLeuThrAlaGluAlaLysAsp 1808
1789GGCGAAGAAAGATACGACCAAAAGC.....AACGG 1817
1029 AspLeuSerGlyLeuAspAlaThrLysSerValLysThrGluIleAsnG 1830
1818 GCGGCTCAACCTT..... 1830
1045 uValThrAsnLeuAspAlaThrPheThrAspAlaGlyThrThrAlaAspG 1862
1831GTTTACAGCGCGCGCGAGAGACCGCCTGCTGCTT 1869
1062 lYThrThrLysIleGluThrProLeuSerAspGluGlnAlaLeu 1878
1870 TCCGCGGGAACAATTTAAACGCAACATCAACGCAAAACGCAACT 1919
1079 GlyAsnGlyAspAsn.....SerAlaGluLe 1087
1920 GTTTTCACGCGGAGACCAACGACGCGCCTACATATTTAAACGACC 1969
1087 uTyrLeu.....ThrAspAsnAlaSerAsnAlaThrAspGlnA 1100
1970 ATTTGCTG...CAAAAGAGGCGATTCCTCGCGGGAATGCTGGGAC 2016
1100 sPAlaSerValGlnLysProGlySerThrSerPheAspLeuIleValAsn 1116
2017 AAC.....GACTGATCAACCGCATTTTAAACGCGAAACTT 2054
1117 GlyGlyGlyIleProAspLysIleSerSerThrThrThrGlyArgIle 1133
2055 CCAATTTAAAGCGGAGACGCGGCTGTTCCCGCAATGTTGCCAAAGTGA 2104
1133 aAsnThrGlnGlyLys..... 1138
2105 AAGCGATTGGCATTTGAGC...AATCAGCCCAAGACGTTTGGTGC 2151
1139 ...GlyThrTyrThrPheSerGlyThrTyrProAlaIleValAspGlyThr 1154
2152GCACCGCATCAAGCCACACATCTGACACGTTGGAGCTG 2192
1155 TyrThrAspAlaGlnGlyLysLysHisAspLeuAsnThrThrTyrAsp... 1170

2193 GACGGTCTGCAAAATTGTGTGCAAAAAACATTACCGAGTAAGTGA 2242
1171 ...AlaIatThrAsnSerPheThrAlaSerMet..... 1180
2243 TTGCTTATGACACAAACGACATGACGGCAATGTCATTT...GCC 2289
1181 ...ProValThrAsnAlaAspTyrAlaAlaGlnValAspLeuTyrAla 1195
2290 GATCAGCGTCATTAAATCTCACAGGCGTTGCCAACATCACTACGCGCATCT 2339
1196 AspTyrAlaHisThrGlnLeuLeuHisPheAspThrLysValArgLe 1212
2340 TAGTGCAAAT...GGCGATACAGTTATACAGTACGCCACAGCCCA 2383
1212 uMetAlaProThrPheThrAspLeuLysPheAsnAsnGlySerAspLeu 1229
2384 CCCAAACGCGCAACCTTAGCTGCTGGCAATGCCCAAGCAACATTAAAT 2433
1229 hSerGlnAlaThrIleLysValThrGlyThrValSerAla....Asp 1243
2434 CAAGCCACATTAAAC...GCCAACATCGGCTCGGCAATGCTTCATT 2480
1244 ThrLysThrValAsnValGlyHisThrValAlaAlaLeuAspAlaGlnH 1260
2481 TAATCTAAGC...GACCAAGCGCTGACAAACGCGCAGTCTGACGCTTCCG 2527
1260 hHisPheSerValAspValProValAsnTyrGlyAspAsnThrIle... 1275
2528 GCACAGCTAAGCAACGTAACCATTCGGACATCAAGGTAAAGTCTCC 2577
1276 ...LysValThrAlaThrAspLysAspLysValAsnThrThr 1287
2578 CTAGCCGATAAGCAGTA..... 2585
1288 ThrGlnGlnLysThrIleThrSerSerTyrAspProAspMetLeuLys 1304
2596 ...TTCCAAAT 2602
1304 sSerValThrPheAspGlnGlyValLysPheGlyThrAsnLysPheAsn 1321
2603 TTGAAACACGCGCTT.....ACC 2622
1321 LathSerAlaLysPheTyrAspProLysThrGlyIleAlaThrIleThr 1337
2623 GCACAAATCAGCGCGGCAAGATAGC.....GCAT 2654
1338 GlnLysValLysHisProThrThrThrLeuGlnValAspGlyLysGlnH 1354
2655 ACACCTTAAAGACAGC.....GAATGACGCTGCGCTCAGCAGCG 2695
1354 eProLysAspAspLeuThrPheSerPheThrLeuAspLeuGlyThr. 1370
2696 AATTAGC..... 2703
1371 ...LeuGlnLysProPheGlyValValGlyAspThrThrGlnAsn 1386
2704 ...AA 2705
1387 LysThrPheGlnGlnAlaLeuSerPheIleLeuAspAlaValAlaProth 1403
2706 TTAAACCTTGACAAACGACCATTAACATTCGCTATGCGCACG 2755
1403 rLeuSerLeuAspSerSerThrAspAla....ProValTyrThrAsn 1418
2756 ATGCGGCAAGGGCGCAACCGGACGTGCGAAGATGCGCGCGCGCGCGT 2805
1418 sProAsnPheGlnIleThrGlyThrAlaThrAspAsnAlaGlnTyrLeu 1434
2806 TCGCGCGCTGCGCGCGCTCCCTATTA..... 2832
1435 SerLeuSerIleAsnGlySerSerValAlaSerGlnTyrGlnAspLeu 1451
2832 2832

1451 nIleAsnSerGlyLysProGlnHisMetAlaIleAspLeuProValLysL 1468
2832 2832
1468 euleuGlnGlyLysAsnValLeuThrValAlaValThrAspSerGlnAsp 1484
2832 2832
1485 AsnThrThrThrLysAsnIleThrValTyrTyrGlnProLysLysThrIle 1501
2833 ...TTCGTTACCGGCCCACTTCGTTGAGAAATCCGTTTCACA 2872
1501 uAlaAlaProThrValThrProSerThrThrGlnProAlaGlnThrValT 1518
2873 CGCTGACGGTAAACGCGCAATTCAGCGCTCAG..... 2904
1518 hLeuThrAlaAsnAlaAlaAlaAlaThrGlyGlnThrValGlnTyrSerAla 1534
2904 2904
1535 AspGlyGlyLysThrTyrGlnAspValProAlaAlaGlyValThrIleTh 1551
2905 ...GAAACATTCGCTTATGTCG...GAACCTCGGCTACCGCA 2944
1551 rAlaAsnGlyThrPheLysPheLysSerThrAspLeuTyrGlyAsnGln 1568
2945 GCGACAAATTCAGCTGCGGAAAGTTCCGAAGCAGCTTACACCTTGCG 2994
1568 eProAlaValAspTyr.....ValVal 1575
2995 GTCAACATATACGCGCACAGAACCTTCGACATTCAGAGGTAGT 3044
1576 ThrAsnIleLysAlaAspProAla.....GlnLeuGlnAlaAl 1589
3045 GGAAGAAAGAACAAACACCGCTGCGAAACCTTAATTTCACCTGTC 3094
1589 a.LysGlnGlnLeuThrAsn.....LeuIleAlaSerAla 1600
3095 AAACGACACAGCTGATGCGCGCGCTGCGTTACCACTCATCCGCAAA 3144
1601 LysThr.....LeuSerAlaSerGlyLysTyrAspAspAlaThrTh 1614
3145 GACGCGAGTTCGCGCTGCATATCCGTCGAAGAACAAGACTTTCGA 3194
1614 rThrAlaLeuAlaAla..... 1619
3195 CAAACTCGCAGACGAGAAACCAAAACAGCGCGGAAAGAACACGCGC 3244
1620 ...AlaThrGlnLysAlaGlnThrAlaLeuAspGlnThrAsn..A 1633
3245 AAGCCTTGACCGCTGATTCGCGCGCGCGGCGATGCTCGAAAGACA 3294
1633 lAserValAspSerLeuThrGlyAlaAsnTyrAspLeuGlnThrAlaIle 1649
3295 GAAAGCCTTCGCGCAACCGCGCGCGGAGGAGGCGGAAATGTCGCGAT 3344
1650 AsnGlnLeuAla..... 1653
3345 TATCAGCGGAGGAGAAAGAAAAAGCGGTGACGGGATTAAGACACCG 3394
1654AlaLysLeuProAlaAspLysLysThS 1663
3395 CCTTGCGCAACACGCGGAAGCG..... 3417
1663 eThrLeuAsnGlnLeuGlnSerValLysAspAlaLeuGlyThrAspLeu 1679
3418 ...GAAACCGCGCGCTACG...ACGCGCTTCCCGCGCGCGCGCG 3458
1680 GlnAsnGlnThrAspProSerThrGlyLysThrPheThrAlaAla.... 1694
3459 GCGCGCGCGGATTTGCGCACTGCAACCCCAACCCGACGCCCAACCGC 3508
11 11

```

1695 .....LeuAspAspLeuValAlaGlnAlaGlnAlaGlyThrng 1707
3509 AGCGGACGCTGATCAGCCGTTATGCCAATAGCGGTTGAGTAATTTCC 3558
1707 LthrAsp.....AspLndleuGln 1713
3559 GCGACGCTCAACAGCGTTTCCCGGTACAGAGCAATTAGACCGCGTATT 3608
1714 AlaThrLeuAlaLysile.....LeuAspGluValle 1724
3609 TCCCGGAG.....ACCGCGGACAGCGG..... 3631
1724 uLaLysLeuAlaGlnGlyLeuLysAlaAlaThrProAlaGluValGly 1740
3632 .....TTTGGCAAGCGCGATCCGG.....ACACC 3657
1741 AsnAlaLysAspAlaAlaThrGlyLysThrPrpAlaAlaAlaAs 1757
3658 AAACACTACCGTTCCAGATTTCCGCGCTACCGCAACACGACCT 3707
1757 pThrLeuThrSerGlyGlnAlaSerAlaAlaSerAspLysLeuAlaH 1774
3708 GCGCAAAATCGGTATGCAAAAACTCGGAGCGGCGGCGGATCC 3757
1774 lSLeuGlnAlaLeuGlnSerLeuThrLysValAlaAlaAlaValGlu 1790
3758 TGTTCGCAACAGCGGCAACCAACCTTCGACGAGCGCATCGGCAAC 3807
1791 AlaAlaLysThrValGly...LysGlyAspGlyThrThrGlyThrSerAs 1806
3808 TCGG.....CACGCTTCCGACGCGGCGGTTTCGCGCAATACGG 3848
1806 pLysGlyGlyGlyGlyThrPro...AlaProAlaProGlyAspThrG 1822
3849 CATGCAAGGTTCTACATCGGCAATCGCGGCGGCGGTTTACGACGG 3898
1822 LysLysAspLysGlyAspGlyLysSer..... 1830
3899 GCAGCCTTCAGACGCGGAGGCAAAATCCGCGCGCTGCTGCAT 3948
1831 .....GlnProSerSerGlyGly.AsnLLeuProThrLysProAla 1844
3949 TAGGCGCTTACGACGATACCGCGCGGTTTCGCGGATTCGCGATCGA 3998
1844 hThrThrSerThrThr.....ThrsAsp 1852
3999 ACCGACATCGGCGCAACGCGTATTGCTCCAAAAAGCGGATTACCG 4046
1853 ThrThrAspAlaGlnGlyGlnLeuThrSerGlyLysGlyAlaLeuPro 1868
seq_name: SwissProt_40:120K_RICRI
seq_documentation_block:
ID 120K_RICRI STANDARD; PRT; 1300 AA.
AC P14914;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 120 kDa surface-exposed protein.
GN P120.
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=783;
RN 11;
RP SEQUENCE FROM N.A.
RX STRAIN-R.
RX MEDLINE=90136087; PubMed=2515418;
RA Gilmore R.D. Jr., Jostle N., McDonald G.A.;
RT Cloning, expression and sequence analysis of the gene encoding the
RT 120 kD surface-exposed protein of Rickettsia rickettsii.
RL Mol. Microbiol. 3:1579-1586(1989).
CC -!- FUNCTION: MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS

```

```

CC RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC -!- MISCELLANEOUS: ITS C-TERMINUS POTENTIALLY MAY BEAR THE EPTIOPES
CC CONFERRING ANTIGENICITY TO THE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X16353; CAA34402.1;
DR PIR: S07575; S07575.
DR InterPro: IPR003858; rompa_ompb.
DR Pfam: PF02708; rompa_ompb; 1.
KW Antigen; Glycoprotein; Cell wall; S-layer.
FT CARBOHYD 66 66 N-LINKED (POTENTIAL).
FT CARBOHYD 86 86 N-LINKED (POTENTIAL).
FT CARBOHYD 103 103 N-LINKED (POTENTIAL).
FT CARBOHYD 147 147 N-LINKED (POTENTIAL).
FT CARBOHYD 268 268 N-LINKED (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (POTENTIAL).
FT CARBOHYD 415 415 N-LINKED (POTENTIAL).
FT CARBOHYD 424 424 N-LINKED (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (POTENTIAL).
FT CARBOHYD 436 436 N-LINKED (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (POTENTIAL).
FT CARBOHYD 547 547 N-LINKED (POTENTIAL).
FT CARBOHYD 593 593 N-LINKED (POTENTIAL).
FT CARBOHYD 655 655 N-LINKED (POTENTIAL).
FT CARBOHYD 698 698 N-LINKED (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (POTENTIAL).
FT CARBOHYD 799 799 N-LINKED (POTENTIAL).
FT CARBOHYD 800 800 N-LINKED (POTENTIAL).
FT CARBOHYD 826 826 N-LINKED (POTENTIAL).
FT CARBOHYD 844 844 N-LINKED (POTENTIAL).
FT CARBOHYD 861 861 N-LINKED (POTENTIAL).
FT CARBOHYD 879 879 N-LINKED (POTENTIAL).
FT CARBOHYD 920 920 N-LINKED (POTENTIAL).
FT CARBOHYD 926 926 N-LINKED (POTENTIAL).
FT CARBOHYD 1116 1116 N-LINKED (POTENTIAL).
FT CARBOHYD 1128 1128 N-LINKED (POTENTIAL).
FT CARBOHYD 1140 1140 N-LINKED (POTENTIAL).
FT CARBOHYD 1146 1146 N-LINKED (POTENTIAL).
FT CARBOHYD 1211 1211 N-LINKED (POTENTIAL).
SQ SEQUENCE 1300 AA; 132801 MW; E09E52C3F647243D CRC64;

alignment_scores:
Quality: 203.50 Length: 1309
Ratio: 0.344 Gaps: 55
Percent Similarity: 45.225 Percent Identity: 19.099

alignment_block:
US-09-303-518d-649 x 120K_RICRI ..
Align seg 1/1 to: 120K_RICRI from: 1 to: 1300
1009 AATGCGACAGGAAATGCAATGCAACATGACACATCTCTGCGCTAA 1058
|||||.....| 16
7 AsnAlaThrGlyGlnValAlaSerPheArgHis..... 16
1059 TAGATTAAACACGACGATGTTTAAATGTTCTTATCGAGA 1108
|||.....| 20
17 .....lleValAspVal..... 20

```

1109 CAGCAGAGAACTGTTATCATCTGCTGTCACAGCTATACGA 1158
21GlyAlaAspGlyThrThrAlaPheLys 29
1159 CCC.....AGACTGAATPANGAGCAAAATATTCCTTAT 1193
30 ThrAlaAlaSerLysValThrIleThrGlnAspSerAsnPheGlyAsnTh 46
1194 TCACGAGGAAAA.....GGCGAATTCGATAC 1219
46 rAspPheGlyAsnLeuAlaIleGlnIleLysValProAsnAlaIleThrL 63
1220 TTACGACGAAACATC.....AATCAAGT.....GCTGGA 1248
63 eutThrLysAsnPheThrGlyAspAlaSerAsnProGlyAsnThrAlaGly 79
1249 GGATATATATTTCCAAAGAGATTTACGCTCGCTGAAATTAACGAAC 1298
80 ValIleThrPheAspAlaAsnGlyThrLeuGlnSerAlaSerAlaAsp.. 95
1299 TTGGCAAGCGCGGCGCTCATATCAGTGAAGACAGTACCGTTACTTGA 1348
96AlaAsnValAlaValThrAsnAsnIleThrAlaIle...G 108
1349 AAGTAAACGGCGCTGCAAAACGCGCTGTCGAAATGCGCAAGGCG 1398
108 lUnIAserGlyAlaGlyValAlaGlnLeuSer.....GlyThr 120
1399 CTGCAGCTTCACGCAAGGCGGAAACCAAGCGTCGATC...AGCGTGG 1445
121 HisAlaAlaGlnLeuAlaGlyLeuGlyAsnAlaGlySerIlePheLysLeuAl 137
1446 CGACGCTACAGTCAATTTGGATCAGCAGACGATTAAGCAAAANAC 1495
137 aAspGlyThrValIle.....AsnGlyLysVala 147
1496 AAGCTTATGAAATCGCTGTCAGCGGCGAGGCGTACGCGCAACTG 1545
147 snGlnThrAlaLeuValGlyGlyAlaLeuAlaIleGlyThrIleThrLeu 163
1546 AATCGCATATATCAGTCAACCCCGAC..... 1572
164 AspGlySerAlaThrIleThrGlyAspIleGlyAsnAlaGlyGlyAlaAl 180
1572 1572
180 aAlaLeuGlnArGlyIleThrLeuAlaAsnAlaLysLysThrLeuThrL 197
1573AATCTATTCGCTTTCGCGGCGAGCGTTCGATTA... 1611
197 euGlyGlyAlaAsnIleIleGlyAlaGlyGlyThrIleAspLeuGln 213
1612 ..AAGGGCATTCGCTTTCGCTTTCGATTCATAAT..... 1647
214 AlaAsnGlyGlyThrIleLysLeuThrSerThrGlnAsnIleVala 230
1648ACCGATGAAGGCGCGATGATTCACAC 1675
230 lAspPheAspLeuAlaIleAlaThrAspGlnThrGlyValValAspAla 247
1676 ACAATCAAGACAAAGATCACCGCTTACCATTAAGCAATTAAT 1725
247 eSerLeuThrAsnAlaGlnThrLeuThrIleAsnGlyLys.....Ile 261
1726 GCTACACCGCAATTAACAC.....ACCTTGATAG 1757
262 GlyThrIleGlyAlaAsnAsnLysThrLeuGlnPheAsnIleGlyLys 278
1758 CAAAAAAGAAATTCCTACACGCTTGGTTGGCAGAAAGAT..... 1800
278 rSerLysThrValLeuSerAsnGlyAsnValAlaIleAsnGlnLeuValI 295
1800 1800

295 lGlyAsnAspGlyAlaValGlnPheAlaHisAspThrIleLeuThr 311
1801 ..ACGACCAAAACGAAACGGCGGCTCAACCTGTTACACGCGCGC 1847
312 ArgThrThrAsnAlaAlaGlyGlnGlyLysIleIlePheAsnProVala 328
1848 AAGAGACCGCACCTGCTGCTTCGCGGAGACAAATTAACGGCAACA 1897
328 lAsnGlnLysThrLeuAlaAlaGlyThrAsnLeuGly..... 341
1898 TCACGCAACAAACGCGCAACTGTTTTCAGCGGACAGACCAACGCGCAC 1947
342 ..SerAlaThrAsn..... 345
1948 GCCTACATTCATTTAAAGCATTGTCGCAAAAAGAGGCGATTCCTCG 1997
346ProLe 347
1998 CGGAAATTCGTGTGGCAACGACTGATCACCGCATTTAAACGG 2047
347 uAlaGlnIle..... 350
2048 AAACTTCCAAATTAAGCGGAGCGGCGGCTTCCCGCATGTTGCC 2097
351 ..AsnPheGlySerLysGlyValAsnValAspThrValLeuAsnValGly 366
2098 AAGGTAAAGCGGATTCGATTCAGCAATCAGCGCCAAAGATTTTGG 2147
367 Gln.....G1 368
2148 TGTGCAACCGCATCAAGCCACACATCTGTCACAGTTCGACTGG... 2193
368 yAlaAsnLeuTyAlaThrAsnIleThrThrAspAlaAsnValGlyS 385
2194ACGGCTGCAAAATTCGTGCAAAACCAT... 2226
385 ePheValPheAsnAlaGlyGlyThrAsnIleValSerGlyThrValGly 401
2227ACGACGATTAAGT 2240
402 GlyGlnGlnLysAsnLysPheAsnThrValAlaLeuGlnAsnGlyThrTh 418
2241 GATTGCTTCATTCAGTAAGCCGACATCAGCGCAATGTCGATTCGCG 2290
418 rValLysPheLeuGlyAsnAlaThrPheAsnGlnAsnThrIleAla 435
2291 ATCAGCTCATTTAAATCTCACAGGGCTGCCACACTCAACGCAATCTT 2340
435 lAsnSerThrLeuGlnIleGlyGlyAsnTyThrAlaAspCysValaAla 451
2341 AGTCAATATGCGCATCACCTTATACAGTCAGCCACACGCAACCCANA 2390
452 SerAlaAspGly..ThrGlyLeuAlaGlnPheValAsnThrGlyProIle 468
2391 CGGCAACTTATGCTCGTGGCAATGCCCAAGCAACATTTAAT..CAAGC 2439
468 hrValThrLeuAsnLysGlnAlaIleAlaProValAsnAlaLeuLysGlnIle 484
2440 ACATTAACGCAACACATCGGCTTCGCGCAATGCTCATTTAATCAAG 2489
485 ThrValSerGlyProGlyAsnValAlaIleAsnGlnIleGlyAsnAlaG 501
2490 CGACGAC.....GCCGTCAAAACGCGCAGTC 2515
501 yAsnHisHisGlyAlaValThrAspThrIleAlaPheGlnAsnSerSerL 518
2516 TG..... 2517
518 euGlyAlaValAlaPheLeuProArGlyIleProPheAsnAspAlaGly 534
2518 ..ACGCTTCCGCAACGCTAAGCAACGTAAGCCATTCGCACTCAA 2564
..... 2564

535 AsnThrMetProLeuThrThrIleLysSerThrValGlyAsnLysThrAlaIle 551
2565 CGGTAATGTCCTCCAGCCGATTAAGCAGATATTCATTGTAAGACGACG 261
551 SGLY...PheAspValProSerValIleValIleuGlyValAlaSpSerValI 567
2615 GCTTACCGGCAATATCGGGC 2637
567 LeuIaAspGlyGlnValIleLysAspGlnAsnAsnIleValIleuGly 583
2658 ...GGCAAGATATGAGCATTAACACTTAAGAACGACGATGACGATGCGC 2684
584 LeuLysSerAspAsnGlyIleIleValAlaSerLhr ThrLeuTy 598
2685 GTGACGACGAGGAATTAGGCAATTAAACCTTGACACGCCACCATTAAC 2734
598 rAlaGly IleSerThrLeuAsnAsnAsnGlnGlyThrValThrL 613
2735 TCATATCCGCTATTCGCCAGATGCGGACGAGGGCGCAACCGGAGTGGC 2784
613 euserglyGlyVal...ProAsnThrProLysThrValITyrglyLeuGly 628
2785 ACAGATGCGCGCGCGCGCTTGCGCGCTTGCGCGCTTCCTATTATC 2834
629 ThrGlyITleGlyAlaSerLysPheLysGln 638
2835 CGTTACACGCCCAACTTCGTTAGATACCGCTTACACACCGCTACGTTA 2884
639 ...ValThrPheThrThrAspTyrsAsnLeuGlyAsnIleIleAlaThrA 655
2885 AC 2886
655 snAlaThrIleAsnAspGlyValIThrValIThrGlyGlyIleAlaGly 671
2887 GGCAATGAAGAGGTGACGGAC 2909
672 IleGlyPheAspGlyLysIleThrLeuGlySerValAsnGlyAsnLys 688
2910 ATTGGCTTATGTGGGAAGCTTTCGCGCTACCGCAACGCAATTGAAAC 2959
688 nAlaITrPheAlaAspGlyIleLeuSerAsnSerThrSerMetIleGlyT 705
2960 TGGCGGAAGTCCAGACGACTTACACTTGGCGTC 2997
705 hrThrLysAlaAsnAsnGlyThrValIThrTyLeuGlyAsnAlaPheVal 721
2998 AACATACCGGCACGACTGCAAGCCTCGAACATTAACCGT 3041
722 GlyAsnIleGlyAspSerAspIThrProValAlaSerValAlaGpHeThr.. 737
3042 AGTGAAGGAAACACACCAACAAACGCTGTCCGAAACCTTAATTTCACC 3091
738GlySerAspSerGlyAlaGlyLeuGlnLysAsnI 749
3092 TGCAAAACGAACACGTCGATCGCGCGCGTGGCTTACCACTCATCGC 3141
749 IeTySerGlnValIleAspPheGlyThrTyrsAsnLeuGlyITleValAsn 765
3142 AAGAGCGCGAGTTCGCGCTGCAATATTCGCTCAAAAGACAGACTTTC 3191
766 SerAsnIleIle 769
3192 CGACAAACTGCGAAGCGACAGACCAAAAAACAGCGCGAAAGACACAAG 3241
770LeuGlyITleGlyIThrThrAlaIleAsnGlyLysITleAspLeuV 784
3242 CGCAAAAGCCTTGACGGCGCTGATTGCGCGCGCGCGCATGCGTGGAAAG 3291
784 alIThrAsn...ThrLeuThrPheAlaSerGlyIThrSerIThrTrpGlyAsn 799
3292 ACAGAAAGCGTTGCGCAACCGCGCGCGACGAGCGGCGGAAATATGTCG 3341
800 AsnIThrSerITleGlyIThrIThrIThrLeuAlaAsnGly...AsnIleG 815

[illegible]


```

225 sPLysGlyGlyIleTyrSerLysGlu.....Lys 235
2044 GCGGAAACTTCCAAATTAAGGCGGACGGCGGTGTTCCCGCAATGT 2093
      :::::::::::::::::::::
226 AspSerThrLeuAspAlaAsnThrGlyValAlaThrPheLysSerAsnThr 252
2094 TGCCAAAGTGAAGGCGGATGGCAT.....TTGAGCA 2125
      :::::::::::::::::::::
252 rAlaLysThrGlyAlaLerTrpSerSerAspAsnLeuAlaLeuThrG 269
2126 ATCAGGCCACGACGCTTTT.....GGTGTGCGACCG 2157
      :::::::::::::::::::::
269 LysAsnThrGlnValLeuPheGlnGlnLysThrThrGlySerAlaAla 285
2158 CATCAAGCCAC.....ACAATCGT..... 2178
      :::::::::::::::::::::
286 GlnAlaAsnAsnProGlnGlyCysGlyGlyAlaIleCysGlyTyrLeuAl 302
2179 .ACACGTTGCGACTGACGGGCTCTG.....ACAAATTGTGCGAATAA 2221
      :::::::::::::::::::::
302 aThrAlaThrAspLysThrGlyLeuAlaIleSerGlnAsnGlnLysL 319
2222 CCATTACCGAGATAAGTATGATTCCTCA.....TTGACT 2256
      :::::::::::::::::::::
319 eRPhetThrSerAsnThrThrThrAlaAsnGlyAlaIleTyrAlaThr 335
2257 AAGACCGACATCAGCGGCAATGTCATCTTGCC..... 2289
      :::::::::::::::::::::
336 LysCysThrLeuAspGlyAsnThrThrLeuThrPheAspGlnAsnThrAl 352
2290 .....GATCAGCGTC 2299
      :::::::::::::::::::::
352 aThrAlaGlyCysGlyAlaIleTyrThrGlnThrLysAspPheSerL 369
2300 ATTAAATCTCACAGGGCTTGCCACACTCAACGCAATCTTATGCAAT 2349
      :::::::::::::::::::::
369 eulysGlySerThrGlyThrValThrPheSerThrAsnThrAlaLysThr 385
2350 GGGGATACGCGTTATACAGTCAGCCACACAGCCCAAAAGCGGACCT 2399
      :::::::::::::::::::::
386 GlyGlyAlaLeuTyrSer.....LysGlyAsnSe 395
2400 TAGCCTGCTGGCAATGCCCAAGCAACATTT.....AATCAAGCCACAT 2443
      :::::::::::::::::::::
395 rSerLeuThrGlyAsnThrAsnLeuLeuPheSerGlyAsnLysAlaThrG 412
2444 TAAAGGCAACACATCGGCTTCG.....GGCAATGCTTCATTT 2481
      :::::::::::::::::::::
412 LysProSerAsnSerSerAlaAsnGlnGlnGlyCysGlyAlaIleLeu 428
2482 AATCTAAGCGACGACGCC.....GT 2501
      :::::::::::::::::::::
429 SerPheLeuGlnSerAlaSerValSerThrLysGlyLeuTrpIleG 445
2502 ACAAAAGCGGAGTGAAGCTTTCCGCAAGCAAGTGAAGCAAGCAAG 2550
      :::::::::::::::::::::
445 uAspAsnGlnAsnValSerLeuSerGlyAsnThr..AlaThrValSerG 461
2551 .....CATTCGCACTCAACGGTAATGTCTCCCTA 2580
      :::::::::::::::::::::
461 LysGlyAlaIleTyrAlaThrLysCysAlaLeuHisGlyAsnThrLeu 477
2581 GCCGATAGAAGCAGTATTCATTTGAAAGCAGCGCTTACCGACGCAAT 2630
      :::::::::::::::::::::
478 Thr.....PheAspGlyAsn.....ThrAlaGlnL 486
2631 CAGGCGGCGGAGATAGCATTAAGCAAGCAAGCAAGCAAGCAAGCAAG 2680
      :::::::::::::::::::::
486 rAlaGlyGly.....AlaIleTyrThrGlnThrGlnAspPheThrL 500
2681 TGCCGTCAGGCAAGCAATTAAGCAATTAACCTTACCAAGCCACCAT 2730
      :::::::::::::::::::::
500 eulThrGlySerThr.....GlyThrVal 507

```

```

2731 ACACTCATTCGCCCTATGCCAGATGCCGACAGGCGCAACCGGACAG 2780
      :::::::::::::::::::::
508 ThrPheSerThr.....AsnThrAlaLysThrIleG 518
2781 TGCGACAGATCGCGCGCGCGCGCTTGCGCCCGCTGCGCGCGCTTCCCTAT 2830
      :::::::::::::::::::::
518 yAlaLeuHisThrLysGlyAsnThrSerPheThrLysAsnLysAlaLeu 535
2831 TATCGGTTACACCGCCACACTTGGGTAGATCCGCTTCAACACGCTGACG 2880
      :::::::::::::::::::::
535 aL.....PheSerGlyAsnSerAlaThrAlaThrAlaThr 546
2881 GTAAACGGCAAAATGAACGGTCAGGACACTTCCGCTTATATGCGACT 2930
      :::::::::::::::::::::
547 ThrThrThrAspGlnGlnGlyCysGly.GlyAlaIleLeuCysAsnIleS 563
2931 CTTCGGCTACCGGACGACAAA.....TTGAAGCTGGCGGAAAGTTCCGAA 2976
      :::::::::::::::::::::
563 eRglnSerAspIleAlaThrLysSerLeuThrLeuThrGlnLysnGlnSer 579
2977 GGCACTTACACCTTGGCGGTCACCAATACCGCAACGACGCAAGCTGCAAGC.. 3024
      :::::::::::::::::::::
580 LeuSerPhe.....IleAsnAsnThrAlaLysArgSerGlyGly 593
3025 ....CTCGAACAAWTGACGCTAGTGAAGCAAGCAACAAACACCGCTGT 3070
      :::::::::::::::::::::
593 yIleTyrAlaProLysCysValIleSerGly.....S 604
3071 CGGAAACCTTAATTTCACCTGCAAAACGAAACGACGTGATCCGCGCG 3120
      :::::::::::::::::::::
604 eRglnSerIleAsnPheAspGlyAsnThrAlaGlnThrSerGlyGlyAla 620
3121 TGCGCTTACCAACTCATCCGCAAGCGGCGAGTCCCGCTGCAATATCC 3170
      :::::::::::::::::::::
621 .....IleTyrSerLysAsnLeuSerIleThrAlaAsnGlyPr 633
3171 GGTCAAGACACAAGAGCTTTCGACAAACCTGCGCAAGCGCAAGCAACCA 3220
      :::::::::::::::::::::
633 oValSerPheThrAsnAsnSerGlyGlyGlyAlaIleTyrIleA 650
3221 AACAGCGGAAAGAACAGACGCGCAACGCTTGAGCGGCTGATTTGCG... 3267
      :::::::::::::::::::::
650 lAspSerGlyGlnLeuSerLeuGlnAlaIleAspGlyAspIleThrPhe 666
3268 GCGGCGCGGATGCCGTGAAAGACAGAAAGCTTCCGCAAGCGGCGCG 3317
      :::::::::::::::::::::
667 SerGlyAsnArgAlaThrGlnGlyThrSerThrProAsnSerIleHisLe 683
3318 GCAGCGAGCGGCGGAAATGTCGCAATTAATGACGCGGAGAGAGAGAAA 3367
      :::::::::::::::::::::
683 uGlyAlaIleAlaLysIleThrLysLeuAla.....AlaAla 695
3368 AACGGGTGACGCGGATAAAGACACCGCTTGCGGAAACGCGCGGAAAGC 3417
      :::::::::::::::::::::
696 ProGlnHisThrIleTyrPheTyrAspProIleThrMetGlnAlaProAl 712
3418 GAACCGCGCGCGCTA..... 3433
      :::::::::::::::::::::
712 aSerGlyGlyThrIleGlnLysLeuValIleAsnProValValLysAlaI 729
3434 .....CCACGCGCTCCCGC..... 3448
      :::::::::::::::::::::
729 lValAlaProProGlnProGlnProLysAsnGlyProIleAlaSerValProVal 745
3449 GCGCGCGCGCGCGCGCGCGGATTTCCGCAATGCAACCCCAACCGGAG 3498
      :::::::::::::::::::::
746 ValProValAlaProAl.....AsnProAsnThrG 756
3499 CCCCAACCGGAGCGGACGCTGATGACGCGTATGCCAATAGCGGTTTGA 3548
      :::::::::::::::::::::
756 y..... 756

```

```

3549 TGAATTTCCGCGCAGCTGTTTCCGCGTACAGCAATGAG 3598
757 .....ThrlleAlpheSer..... 761
3599 ACCGCGATTATTCGCGAGACCGCGCAACGCGTTTGACACAGCGCATC 3648
762 SerGlyLysLeuPro.....SerGlnAspAla 771
3649 CGGACACCAACACTACCTT.....CCGAGATTTCGGCGCCCTA 3689
771 rtleProAlaSerThrThrThrLeuAsnGlnLysLleAsnLeuAlaG 788
3690 CCGCCCAACAAACGACCTGCGCAATCGGTATCGAGAAAACCTCGCA 3739
788 lyeIAsn.....ValValLeuLys..... 794
3740 GCGGCGCGCGCATCTCTTTTCGCAACACCGGACCGCAACACCTTC 3789
795 GluGlyAlaThrLeuGlnValTyrSerPheThr.....GlnGlnProAs 809
3790 GACGACGGCAGTCGCACTCGCGCAGCGCTTGCCACGCGCGCTTTTCG 3839
809 pserThrVal.....PheMetAspAlaGlyThrThrL 820
3840 GCAATACGCGCATCGACAGTTCATCGCATCGCGCGCGCGCGGTT 3889
820 euglThrThrThrThrAsnAsnThrAspGlySerLleAspLeuLysAsn 836
3890 TTAGACGCGCGCAGCCTTCAGCGCATCGGCAAAATCGCGCGCGC 3939
837 Leu.....SerValAsnLeu.....AspAlaLeuAspGlyLys.....Arg 848
3940 GTGCTCATTCACGCGCATTCAGCGCATACCGCGCGCTTCGCGCGATT 3989
849 MetLleThrLleAlaValAsnSerThr.....SerGlyLysLe 861
3990 CGGCGCATCGACGCGCATCGCGCAACGCGCTATTCTGTCMAAAAGCG 4039
861 ulysLleSer.....GlyAspLeuLysPheHisAsnGlnLys 875
4040 ATTACGCGCTACGAAACGTCAATATGCGACCCCGCGCTTCATTCAC 4089
875 erPhe.....TyrAspAsn..... 879
4090 CGCTACGCGCGCGCATTCAGCATATTCATTCACACCGCGCAACA 4139
880 .....ProGlyLeuLysAlaAsnLeuAsnLeu..... 888
4140 CATTCATCAGCGCTTATTCAGCTGCTCTATACGATCGCGCTTCG 4189
889 .....ProPheLeuAsnLeuSerSerThr.....SerG 898
4190 GCAAAAGTCCGA.....ACACGCGCTCAATACCGCGCTATGCT 4227
898 lYthrValAsnLeuAspAspPheAsnProLleProSerSerMetAlaAla 914
4228 CAGATTTCGCGC..... 4239
915 ProAspTyrGlyTyrGlnGlySerTyrThrLeuValProLysValGlyAl 931
4240 .....AAAACCGCGCAGTGGGGAATGGGCGCTAAACGCGCAATCA 4279
931 aglYglYlysValThrLeuValAlaGluTyr.....GlnAlaL 944
4280 AAGGTTTCAGC 4290
944 euglYtyrThr 947

```

seq_name: SwissProt_40:P2P_LACLC

seq_documentation_block: PRT; 1902 AA.

ID P2P_LACLC
AC P15293;

```

DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE PII-type proteinase precursor (EC 3.4.21.96) (Lactocoeplin) (Cell wall-
DE associated serine proteinase) (Lp151).
GN
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG Plasmid pUP763.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1359;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=NCDO 763;
RX MEDLINE=89313288; PubMed=2501630;
RA Kiwaki M., Ikemura H., Shimizu-Kadota M., Hirasima A.;
RT "Molecular characterization of a cell wall-associated proteinase gene
RT from Streptococcus lactis NCDO763."
RL Mol. Microbiol. 3:359-369(1989).
CC -1- FUNCTION: PROTEINASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC -1- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
CC specificity, although some substrate preference have been noted,
CC e.g. large hydrophobic residues in the P1 and P4 positions, and
CC pro. in the P2 position. Best known for its action on caseins,
CC although it has been shown to hydrolyse hemoglobin and oxidized
CC insulin B-chain.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC
CC EMBL: X14130; CAA32350.1; -.
CC PIR: S06997; S06997.
CC DR HSSP: P00782; 2SHT.
CC DR MEROPS: S08.019; -.
CC DR InterPro: IPR001899; Gram_pos_anchor.
CC DR InterPro: IPR000209; Peptidase_S8.
CC DR Pfam: PF00746; Gram_pos_anchor, 1.
CC DR Pfam: PF02225; PA; 1.
CC DR PRINTS: PR00723; SUBTILISIN.
CC DR PROSITE: PS00136; SUBTILASE_ASP; 1.
CC DR PROSITE: PS00137; SUBTILASE_HIS; 1.
CC DR PROSITE: PS00138; SUBTILASE_SER; 1.
CC DR PROSITE: PS00433; GRAM_POS_ANCHORING; 1.
CC DR HydroLase: Serine protease; Cell wall; Zymogen; Signal; Plasmid;
CC transmembrane.
CC KW
CC SIGNAL
CC FT 1 33
CC FT PROPEP 187
CC FT CHAIN 188 1902
CC FT DOMAIN 187 1876
CC FT TRANSMEM 1877 1895
CC FT DOMAIN 1896 1902
CC FT ACT_SITE 217 217
CC FT ACT_SITE 281 281
CC FT ACT_SITE 620 620
CC FT DOMAIN 1867 1872
CC FT
CC SEQUENCE 1902 AA; 200139 MW; 4B8DBB64DB8CDF7 CRC64;

```

alignment_scores:

Quality: 199.50
Ratio: 0.360Length: 1275
Gaps: 73

Percent Similarity: 43.451 Percent Identity: 21.098

alignment_block:

US-09-303-518D-649 x P2P_LACLC ..

Align seg 1/1 to: P2P_LACLC from: 1 to: 1902

```

709 GATCATGGTGGTGGACAGTCACTTA..... 735
815 GlySerAspGlySerArgLeuAsnLeuProTyrMetGlyPhePheGlyAs 831
736 ....GGTGTGAAAAAATTAACATPACCCATATGAT.....TTTTTAC 775
831 PTPAsnAspGlyLysIleValAspSerLeuAsnGlyIleThrTyrSerP 848
776 CAACAGAGAGCTCATTTGGCGACAGTGGCTCACCAATGTTATCTATGAT 825
848 roAlaGlyGlyAsnPheGlyThr.....ValProLeu..... 858
826 GCCCAAAAGCAAAAAGTGTAAATTAATGAGGATTCGCAACGCGCAACCC 875
859 .....LeuThrAsn.....LysAsnThrGlyThrGcl 867
876 CTATATAGCA.....AAAGCAATGGCTTCACGCTGGTTCGTA 913
867 nTyrTyrGlyGlyMetValThrAspAlaAspGlyAsnGlnThrVal.... 882
914 AAGATTGGTCTATGATGAATCTTTGCGAGATCCCATTCAGATTC 963
883 .....AspAspGlnAlaIle... 887
964 TAGAACCAAGCTCAAAATGGAAATACCTTTTAACGACGATTAATATGG 1013
888 .....AlaPheSerSerAspLys..... 893
1014 CACAGAAAAATCAATGCCAATGACACCAATCTCTCGGCTATATAGAT 1063
894 .....AsnAlaLeuTyrAsnAspI 900
1064 TAAAAACAGAACCGTTCATATGTTTAATGTTTCTTTATCCGACAGACGA 1113
900 IeSerMetLysTyrTyrLeuLeuAlaGlnIleSer..... 911
1114 AAGAAACCTGTTATCATGCTGACAGTGGTGCAACAGTTATCCAGCCAG 1163
911 ..... 911
1164 ACTGAATATGAGAAAAATATTCCTTT...ATTGACGAAGAAAAAGCG 1210
912 .....AsnValGlnValAspIleLeuAspGlyGlnGly 923
1211 AATTGATACTTACGACCAACATCAATCAGTGGTGGAGATTAATATTC 1260
923 snLysValThrThrLeuSerSerThrAsnArgLysLysThrTyrTyr 939
1261 .....CAAGGAGATTTTACGGTCTCGCGCTGAAATATACGA 1295
940 AsnAlaHisSerGlnGlnTyrIleTyrTyrHisAlaPro..... 952
1296 AACTTGGCAAGGC.....GGGCGGCTTCATATCA 1324
953 .AlaTrpAspGlyThrTyrTyrAspGlnArgAspGlyAsnIleLysThrA 969
1325 GTGAAGACAGTACCGTTACTTGAAGATAAGCGCTGGCAACGAC... 1371
969 IaaAspAspGlySerTyrThrTyrArgIleSerGlyValProGlnGlyGly 985
1372 .....CGCCTGTCCAAAATGGCGCA 1391
986 AspLysArgGlnValAlaPheAspValProPheLysLeuAspSerLysAlaPr 1002
1392 AGGACGCTGCACGTT.....CAAGCCAAAGGGGCAAAACCAAGGCTGCA 1435
::: ||||| ::||| |||||

```

```

1002 oThrValArgHisValAlaLeuSerAlaLysThrGluAsn..... 1015
1436 TCACGGTGGCGGACGCTACACTATTTTGGATGACGACGACGACAT... 1482
1016 .....GlyLysThrGlnTyrTyrLeuThrAlaGlnAlaLysAspAsp 1029
1483 .....AAGGCAAAAAACACGCTTTTGTGTAAT 1511
1030 LeuSerGlyLeuAspAlaThrLysSerValLysThrAlaIleAsnGluVa 1046
1512 CGCCTGGTCAAGCGGAGGGGTACGTCGAATGCAATGCCGATAATACAG 1561
1046 I.....ThrAsnLeuAspAlaThrP 1053
1562 TCACCCCGACAAACCTATTTGGCTTGGCGGCGAGCTTGGATTTA 1611
1053 heThr.....AspAla 1056
1612 AACGGGCAATTCGCTTTCGTTCCACCGTATTCAAAT.....ACGATGA 1655
1057 GlyThrThrAlaAspGlyTyrThrLysIleGlnThrProLeuSerAspGl 1073
1656 AGGGCGCATGATGTTCACACACATCAAGCAAAAGATCCACCGTTACCA 1705
1073 uGlnAlaGlnAlaLeuGly...AsnGlyAspAsnSerAlaGlnLeuTyrL 1089
1706 TTACAGGCATTAAGATATGCTACACCGGCAT..... 1740
1089 eThrAspAsnAlaSerAsnAlaThrAspGlnAspAlaSerValGlnLys 1105
1741 .....ACACACAGCTTGATAGCAAAAAAGAAATGGCTTCACACGTTG 1784
1106 ProGlySerThrSerPheAspLeu.....IleValAsnGlyGlyGcl 1119
1785 GTTTGGCGGAAAGATACGACCAAAACGACGGGGCTCAACCTGTTT 1834
1119 yLleProAspLysIleSerSerThrThrThrGly.....T 1131
1835 ACCAG..... 1839
1131 yGlnAlaAsnThrGlnGlyGlyThrTyrThrPheSerGlyThrTyr 1147
1840 CCCGCCGAGAAAGCGACCGCTGCTTCGCCGGAACAAATTTAA 1889
1148 ProAlaIleValAspGlyThrTyrThrAspAlaGlnGlyLysHisAs 1164
1890 CGGCAACATCACG.....CAACAAACGCGCAACTGTTTTCAGCG 1930
1164 PheAsnThrThrTyrAspAlaIleThrAsnSer.....PheThrAlaAs 1179
1931 GCAGACCAACACCGCAC..... 1947
1179 erMetProValThrAsnAlaAspTyrAlaIleAlaGlnValAspLeuTyrAla 1195
1948 .....GCCATCAATCATTTAAACGACCATTTGGCGCAAAAGAGCGCAT 1991
1196 AspLysAlaHisThrGlnLeuLeuLysHisPheAspThrLysValArgLe 1212
1992 T.....CTTCGCGGGAATTCGTGGGCAACAGACTGATCAACCGCA 2035
1212 uThrAlaProThrPheThrAspLeuLysPheAsnAsnGlySerAspGlnT 1229
2036 CATTTAAAGCGGAAACTTCGCAATTAAGCGGACAGCGGCGGTTC 2085
1229 hSerGlnAla...ThrIleLysValThrGly.....ThrValSer 1241
2086 CGCAATGTTCGCAAAAGTGA...GGCAATTCGATTTGAGCAATACAGC 2132
1242 AlaAspThrLysThrValAsnValGlyAspThrValAlaIleAlaLeuAspAl 1258
2133 CCAAGCAGTTTGTGTG.....GCACCGATCAAAAGCCACACAA 2173
1258 aGlnHisHisPheSerValAspValProValAsnTyrGlnLysPaspThrI 1275

```



```

1761 SerGlyGlnAlaSerAlaAspAlaSerAspSerLeuAlaHisLeuGlnAl 1777
1778 GGTATGCAAAAAAAGCTGCGAGCGGCGCTGCGATCCTGTTTTCGCA 3767
1777 AleuGlnSerLeuSerThrLysValAlaAlaAlaValAlaAlaLysr 1794
3768 CAACCGGACCGAAACACCTTCGACGACGATCGGCAACTCGGCACGCG 3817
1794 hrValGly...LysGlyAspGlyThrThrGlyThrSerAspLysGlyGly 1809
3818 .....TTGCCACG..... 3826
1810 GlyGlnGlyThrProAlaProThrProGlyAspIleGlyLysAspLysGly 1826
3827 .....GCGCCGTTTTCGCGC.....AATACG 3848
1826 YAspGlyGlySerGlnProSerSerGlyGlyAsnIleProThrAsnProA 1843
3849 CATCGACAGGTTCTACATCG 3868
1843 LathrThrThrSerThrSer 1849

```

seq_name: Swissprot_40:N100_YEAST

seq_documentation_block:

```

ID N100_YEAST STANDARD: PRT: 959 AA.
AC 002629;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE Nucleoporin NUP100/NSP100 (Nuclear pore protein NUP100/NSP100).
GN NUP100 OR NSP100 OR YKL068W OR YKL336.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes;
OC NCBI_TaxID:4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93054906; PubMed-138544;
RA Wente S.R., Ront M.P., Blobel G.;
RT "A new family of yeast nuclear pore complex proteins.";
RL J. Cell Biol. 119:705-723(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-5288C;
RX MEDLINE-94378724; PubMed-8091863;
RA Rasmussen S.W.;
RT "Sequence of a 20.7 kb region of yeast chromosome XI includes the
RT nucleoside diphosphate kinase gene, tRNAs for His, Val and Trp in
RT addition to seven ORFs with weak or no significant similarity to
RT known proteins.";
RL Yeast 10:S69-S74(1994).
RN [3]
RP FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.
CC THE PROTEINS DURING NUCLEOCYTOPLASMIC TRANSPORT.
CC -1- SUBCELLULAR LOCATION: Nuclear pore complex.
CC -1- DOMAIN: CONTAINS G-L-F-G REPEATS.
CC -1- SIMILARITY: BELONGS TO THE GLFG FAMILY OF NUCLEOPORINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
DR EMBL: Z15035; CA78753.1;
DR EMBL: X75780; CA53406.1;
DR EMBL: Z28068; CA81905.1;
DR PIR: B44402; B44402.

```

```

DR PIR: S39173; S39173.
DR PIR: S44518; S44518.
DR SGD: S0001551; NUP100.
DR InterPro: IPR004325; Nucleoporin_FG.
DR Pfam: PF03093; Nucleoporin_FG; 24.
KW Nuclear protein; Transport; Repeat.
FT DOMAIN 33 571
FT G-L-F-G.
SQ SEQUENCE 959 AA; 99988 MW; D3985F9901BBA51 CRC64;

```

alignment_scores:

Quality: 195.50 Length: 1077
Ratio: 0.403 Gaps: 52
Percent Similarity: 45.032 Percent Identity: 19.591

alignment_block:

us-09-303-518d-649 x N100_YEAST

Align seq 1/1 to: N100_YEAST from: 1 to: 959

```

352 GCGTATACACAGTTCATTTGTGCGGAA.....GGAAG 386
10 GlyGlySerAsnLeuSerPheGlySerAsnThrSerPheGlyGly 26
387 AAATCCGATCAACATCGTTTACTTATTAATAATGTGAACGATTAAT 436
26 ngInSerGlnGlnProAsnSerLeuPheGlyAsnSerAsnAsnAsn 43
437 ATTAACGACGAGACTAAGCCCATCT...TATGCGGCGATATCAATG 483
43 snSerThrSerAsnAlaGlnSerGlyPheGlyGly..... 55
484 CCGCGTTTGCAATTAATTGTGCAGATGACAGACCTGTTGAATGACAG 533
56 .....PheThrSerAlaAlaGlySerAsnSerAsn 66
534 TTATATGATGGCGCGGAATATATGATCAAAATTAATTAACCTGAC 583
66 rLeuPheGlyAsnAsnAsn.....ThrGlnAsnAsn..... 76
584 TTCGATTGTTGGGACGACGACGATATTCGATCGATGTAAGATGACCC 633
77 .....GlyAlaPheGlyInSerMetGlyAlaThrGlnAsnSerPro 90
634 AATAACCCGGAAGTTCATATCATATTCAGATCGATTCTTGCTCGT 683
91 PheGlySerLeuAsnSerSerAsnAlaSerAsn..... 101
684 TGTGTGCAATACCTTTCACAAATGATCA.....GTGGTG 721
102 .....GlyAsnThrPheGlyGlySerSerMetGlySerPheGlyGly 117
722 GCACATGCACTTA.....GTTAGTGAATAAATTAACATACCCATAT 765
117 snThrAsnAlaAlaPheAsnAsnAsnSerAsnSerThrAsnSerProPhe 133
766 GGTATT..... 771
134 GlyPheAsnLysProAsnThrGlyGlyThrLeuPheGlySerGlnAsnAs 150
772 .....TTACCAAG 781
150 nAsnSerAlaGlyThrSerSerLeuPheGlyGlyGlnSerThrThr 167
782 GAGCGCATTTGGGACAGGCTGCACCAATGTTATATCATGATGCCAA 831
167 hrGlyThrPheGlyAsnThrGlySerSerPheGly..... 178
832 AACCAAAAGTGTTAATTAATGAGGATATTCACAAAGGCAACCCATAT 881
179 .....ThrGlyLeuAsnGlyAsnGlySerAsnIlePhe 189

```

```

882 AGGAAAAAGCAATGCTTCAGCTGCTGTAAGATTGTTCTATGATG 931
    ||| :|||: |||
189 eglYAlaGlyAsnAsnSerGlnSerAsnThrThr.....GlyS 202
    :|||: |||: |||
932 AAATCTTGGTGGAGATACCATTCAGTATTCACGACCGCTCAAAAT 981
    :|||: |||: |||
202 erLeuPheGlyAsnGlnGlnSerSerAlaPheGlyThrAsnAsnGlnGln 218
    :|||: |||: |||
982 GGGAAATACCTCTTTTAACGACGATATATATGACAGGAAAAATCAATGC 1031
    |||: |||: |||
219 GlySerLeu...PheGlyGlnGlnSerGlnAsnThrAsn.....AsnAl 232
    :|||: |||: |||
1032 CAACATGACACAAATCTCTGCTTAATTAATTAACGACGAGCTTC 1081
    |||: |||: |||
232 APheGlyAsnGlnAsnGlnLeuGlyGly..... 241
    :|||: |||: |||
1082 AATTCTTAATGTTTCTTATTCGAGACGACGACGACGACCTGTTATCAT 1131
    |||: |||: |||
242 .....SerSerPheGlySerLysProValGlySer 251
    :|||: |||: |||
1132 GCTGCA.....GCTGCTGTCACAGATTATGACCCGACACTGAATATGG 1175
    :|||: |||: |||
252 GlySerLeuPheGlyGlnSerAsnAsnThrLeuGlyAsnThrThrAsnAs 268
    :|||: |||: |||
1176 AGAAAAATATTCCTTTATTGACGAAAGGAGGAGGAAATGCTACTTACCA 1225
    |||: |||: |||
268 nATGAsnGlyLeuPhe.....GlyGlnMet.....A 277
    :|||: |||: |||
1226 GCAACATCAATCAAGGTGCT.....GGAGATTATATTTCCAGAGAT 1269
    :|||: |||: |||
277 snSerSerAsnGlnGlySerSerAsnSerGlyLeuPheGlyGlnAsnSer 293
    :|||: |||: |||
1270 TTACAGCTTCGCT.....GAAATTAACGAAACTTGGCA 1304
    :|||: |||: |||
294 MetAsnSerSerThrGlnGlyValPheGlyGlnAsnAsnGlnMetG1 310
    :|||: |||: |||
1305 AGCGCGCGGCTTCATATCATGTAAGACAGTACCTTACTTGAAAGTAA 1354
    |||: |||: |||
310 n.....Tlea 312
    :|||: |||: |||
1355 AGGCGGTGGCAACGACCGCTGTCCAAAATCGCAGAA...GGCAGCGTG 1401
    |||: |||: |||
312 snGly.....AsnAsnAsnAsnSerLeuPheGlyLysAlaAsnThrPhe 326
    :|||: |||: |||
1402 CACGTTCAACCCAAAGGGGAAAAACAAGCTGATCAGCTGGCGGACGG 1451
    :|||: |||: |||
327 SerAsnSerAlaSerGlyGlyLeuPheGlyGlnAsnAsnGlnGlnGln1 343
    :|||: |||: |||
1452 TACAGTATTTTGGATTCAGCAGCAGCAGCATTAAGCAAAAAACAAGCT 1501
    :|||: |||: |||
343 ySerGlyLeuPheGlyGlnAsnSerGlnThrSerGlySerSerGlyLeuP 360
    :|||: |||: |||
1502 TTACTGAATCGGCTTGTGTCAGCGGACGAGGCTGCACTGAATGCC 1551
    |||: |||: |||
360 heGlyGln..... 362
    :|||: |||: |||
1552 GATATATCATTCACCCCGACAAACTATTTGGCTTTGCGGCGGAGAG 1601
    :|||: |||: |||
363 AsnAsnGlnLysGlnProAsn..... 369
    :|||: |||: |||
1602 TTTCGATTAAAGCGCATTCGCTTCCACCGATATCAAAATACGG 1651
    :|||: |||: |||
370 .....ThrPheThrGln...SerAsnThrG 377
    :|||: |||: |||
1652 ATGAAGGGCGCATGATTGTCAACCAATCAAGCAAAAGATCCACCGTT 1701
    |||: |||: |||
377 LyIleGlyLeuPheGlyGlnAsnAsnAsnGlnGlnGlnGlnGlnGln 392
    :|||: |||: |||
1702 ACCATTACAGCAATAAAGATATTGCTACAGCGGCAATAAACAAGACTT 1751
    :|||: |||: |||
393 GlyLeuPheGlyAlaLysProAlaGlyThrThrGlySer..... 405
    :|||: |||: |||
1752 GGATGCAAAAAAGAAATTGCTACACGCTTGGCTTGGCGAGAAAGATA 1801
    :|||: |||: |||
406 .....LeuPheGlyGlyAsnSers 412
    :|||: |||: |||
1802 CGACCAAAAGACAGCGCGGCTCAACCTGTTTACAGCCCGCCAGAA 1851
    :|||: |||: |||
412 erThrGlnProAsnSerLeuPheGlyThrThrAsnValProThrSerAsn 428
    :|||: |||: |||
1852 GACCGACCCCTGCTGCTTCCGCGACACAATTTAAACGGCAATCAC 1901
    :|||: |||: |||
429 ThrGlnSer.....GlnGlnGlyAsnSerLeuPheGlyAlaThrLy 442
    :|||: |||: |||
1902 GCAACACAAAGCGCAACCTGTTTTCAGCGGACGACACACGACGCT 1951
    :|||: |||: |||
442 sleuThrAsn.....MetProPheGlyGlyAsnProThrAlaAsnGlnS 457
    :|||: |||: |||
1952 ACAATCATTTAAAGCATTGCTGTCGCAAAAAGAGGCAATTCCTCCGG 2001
    :|||: |||: |||
457 erGlySerGlyAsnSerLeuPheGlyThrLysProAlaSerThrThrLy 473
    :|||: |||: |||
2002 GAAATCGTGGGACAAACGACTGATACCGGACGACATTTAAACGGGAAA 2051
    :|||: |||: |||
474 SerLeu...PheGlyAsnAsnThrAlaSerThrThrValProSerThrAs 489
    :|||: |||: |||
2052 CTTCCAATTTAAAGCGGACAGCGGCTGTTTCCCGCAATGTTGCCAAG 2101
    :|||: |||: |||
489 n...GlyLeuPheGlyAsnAsnAlaAsnAsnSerThrSerThrThrAsn 505
    :|||: |||: |||
2102 TGAAAGCGGATTGGCATTGAGCAATCAGCCCAAGCAGATTTTGGTGTG 2151
    :|||: |||: |||
505 hr.....GlyLeuPheGlyAla 510
    :|||: |||: |||
2152 GCACCGCATCAAGCCACACATCTGTACAGCTTCGAGACGAGCGGTCT 2201
    :|||: |||: |||
511 LysProAspSerGlnSerLysProAlaLeuGlyGlyGlyLeuPheGlyAs 527
    :|||: |||: |||
2202 GACAAATTCGTGCAAAAACCATTAACGAGATTAAGTATTCCTCAT 2251
    :|||: |||: |||
527 nSerAsnSerAsnSerSerThrIleGlyGlnAsnLysProValPheGlyG 544
    :|||: |||: |||
2252 TGACATGACCGCATCAAGCGGCAATGTCATCTTCCGATCAGCTCAT 2301
    :|||: |||: |||
544 LyThrThrGlnAsn..... 548
    :|||: |||: |||
2302 TTAATCTCAAGGCTTCCACACTCAACGCAATTTAGTCAAAATGG 2351
    :|||: |||: |||
549 .....ThrGlyLeuPheGlyAlaThrGlyThr..... 557
    :|||: |||: |||
2352 CGATACAGCTTATACAGTACGACCAAGCCACCAAAACGGCAACCTTA 2401
    :|||: |||: |||
558 .....AsnSers 560
    :|||: |||: |||
2402 GCCTCGTGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAAGGC 2451
    :|||: |||: |||
560 erAlaValGlySerThrGlyLysLeuPheGlyGln.....AsnAsn 573
    :|||: |||: |||
2452 AACCATCGGCTTCGGCAATGCTTCATTTAATCTTAAGCGACCGCGCT 2501
    :|||: |||: |||
574 AsnThrLeuAsnValGlyThrGlnAsnValProProValAsnAsnThrTh 590
    :|||: |||: |||
2502 ACAAAAC.....GGCAGTGTAGCGCTTCCGCAACGCTAAGGCA 2542
    :|||: |||: |||
590 rGlnAsnAlaLeuLeuGlyThrThrAlaValProSerLeuGlnGlnAlaP 607
    :|||: |||: |||
2543 ACGTAAAGCATTCGCACTCAACGGTAATCTCCCTAGCCGATAGCA 2592
    :|||: |||: |||
607 roValThrAsnGlnGlnLeuPheSerLysIleSerIleProAsnSerIle 623
    :|||: |||: |||
2593 GTATTTCATTTTGAAGCAGCGGCTTACGAGCAAAATCAGCGCGGCAA 2642
    :|||: |||: |||
623 ..... 623
    :|||: |||: |||
2643 GGATACGCGCATTAACCTTAAAGACAGCGAATGAGCGCTCCGTCAGCA 2692
    :|||: |||: |||

```

```

624 .....ThrasProValLysa 629
2693 CGGAATTGCAATTAAACCTTGAC.....AACGCCACCATTTACA 2733
      |||.....|||
629 lAtHrThSerLysValAlaAspMetLysArgAsnSerSerLeuThr 645
      |||.....|||
2734 CTCATTCGCGCTATCGCCACGATGCGGAGGCGCAACCGCGAGTGC 2783
      |||.....|||
646 .....SerAlaLysArgLeu..... 650
2784 GACAGATGCGCGCGCGCGCTTGCGCGCTTGCGCGCTTCCCTATTAT 2833
      |||.....|||
651 .....AlaProLys.....P 654
2834 CGGTACACCGCCCACTTCGTAGATCCGTTTCAACAGCTGACGGTA 2883
      |||.....|||
654 rOLEuPheAlaProSerSer..... 660
2884 AACGCAATTTGAACGGTCAACATTCGCTTATGTCGCAACTTT 2933
      |||.....|||
661 .....AsnGlyAspAlaLysPheGlnLysTyrPoliLys..... 671
2934 CGGCTACCGGACGCAAAATTGAAGCTGGCGAAGTTCCGAGGACATT 2983
      |||.....|||
672 .....ThLeuGluArgSerAspArgLys..... 681
2984 ACACCTTGGCGGTCAACATTCGCGGACGACCTGCAAC..... 3024
      |||.....|||
681 eRTHr.....SerAsnSerLeThrAspProGluSerSerTyrLeu 694
3025 .....CTGCAACAATT 3035
695 AsnSerAsnAspLeuPheAspProAspArgTyrTyrLeuLysHisLe 711
3036 GAGCGTGTGGAAGAAACACAAACCGCTGTCGAAACCTTAATT 3085
      |||.....|||
711 uValIle.....LysAsnAsnLys.....AsnLeuAsnV 721
3086 TCACCTCTGCAAAAGCAACACGTCGCGCGGCGGCGGTTCACCACTC 3135
      |||.....|||
721 aIle.....AsnHisAsnAspGluAlaSerLysValLysLeu 734
3136 ATCCGCAAGACGCGGCTTCGCTGCATTAATCCGTCACAAACAGA 3185
      |||.....|||
735 Val.....ThrPheThrThrGluSerAlaSerLysAspAspG 747
3186 GCTTTCGACAACTGCGCAGGACGCAAGCCCAAAACAGCGGAAAAAG 3235
      |||.....|||
747 nAlaSerSerSerIleAla.....AlaSerLysLeuThrGluLysA 761
3236 ACAACGCGCAAGCCTTGACGCGCTGATTCGCGCGGCGGATGCCGTC 3285
      |||.....|||
761 lAlHisSerProGlnThrAspLeu.....LysAsp..... 770
3286 GAAAGACAGAAAGCTGTCGCGCAGCGCGCGGACGCGGAGAAA 3335
      |||.....|||
771 AspHisAspGluSerThrProAspProGlnSerLysSerProAsnGly 787
3336 TGTCGCAATTATGACGCGGAGGAGAAAGAA 3366
      |||.....|||
787 rThrSerIleProMetIleGluAsnGluLys 797
seq_name: SwissProt_40: HXA3_HAEIN
seq_documentation_block:
ID HXA3_HAEIN STANDARD; PRT; 917 AA.
AC P45355;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heme/hemopexin-binding protein precursor (Heme:hemopexin utilization
  protein A).
GN HXA.

```

```

OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxId=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NH1 N182;
RA MEDLINE=95115556; PubMed=7815944;
RA Cope L.D., Thomas S.F., Latimer J.L., Slaughter C.A.,
RA Mueller-Eberhard U., Hansen E.J.,
RT "The 100 kDa heme:hemopexin-binding protein of Haemophilus
RU influenzae: structure and localization."
RU Mol. Microbiol. 13:863-873(1994).
CC -!- FUNCTION: BINDS HEME/HEMOPEXIN COMPLEXES.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -----
CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC or send an email to license@sib.ch).
CC -----
DR EMBL: 008349; AAA74139.1; -
KW Transport; Signal; Repeat.
FT SIGNAL 1 21
FT CHAIN 22 917 BY SIMILARITY.
FT DOMAIN 24 38 HEME/HEMOPEXIN-BINDING PROTEIN.
FT REPEAT 24 28 3 X 5 AA TANDEN REPEATS.
FT REPEAT 29 33 1-1.
FT REPEAT 34 38 1-2.
FT DOMAIN 111 668 6 X 6 AA APPROXIMATE REPEATS.
FT REPEAT 111 116 2-1.
FT REPEAT 111 116 2-1.
FT REPEAT 203 208 2-2.
FT REPEAT 277 282 2-3.
FT REPEAT 399 404 2-4.
FT REPEAT 624 629 2-5.
FT REPEAT 663 668 2-6.
FT DOMAIN 159 170 2 X 6 AA APPROXIMATE TANDEN REPEATS.
FT REPEAT 159 164 3-1.
FT REPEAT 165 170 3-2.
SQ SEQUENCE 917 AA; 100148 MW; 7FF39BB8C046539D CRC64;

alignment_scores:
Quality: 195.00 Length: 1095
Ratio: 0.390 Gaps: 60
Percent Similarity: 45.662 Percent Identity: 20.183

Alignment_block:
US-09-303-518D-649 x HXA3_HAEIN ..
Align seg 1/1 to: HXA3_HAEIN from: 1 to: 917
442 GCAGGAGCTAAAGCCATCTTATGCGGC.....GATATCATATGCC 485
      |||.....|||
21 AlaserThrGlnGlyLeuProGlnGlnGlyLeuProGlnGlnGlyLeuPr 37
486 GCGTTTGATTAATTGTGCACAGATGCAAGACCTGTTGAATGACGATT 535
      |||.....|||
37 GlnGlnLysPlyValValAlaValGlyGlnAlaThrPheAspLysThr.... 52
536 ATATGATGCGCGGCAATATATGATGATTAATTAATTAATTAATTAAT 585
      |||.....|||
53 .....lAlaAspLysMetThrIleAsnGlnThrSer.....AspLysVal 66
586 CGTATTGGGCGGACGACGCAATATGCGCATCTGATGAGATGACGCCAA 635
      |||.....|||
67 GlnIleAsn.....TrpHisSer...PheAspIleGly 77
636 TAACCGCAAGTTTCATATCATATTCGAAGT.....GGGTATT 673
      |||.....|||

```



```
2096 CCAAGTGAAGCGATTGGCATTTGACATTCAGCCGACGACGATTTT 2145
      :::::|
      620 LeuSlnLeuLys..... 623
2146 GGTGTGCGACCGCATCAAGCCACACATCTGTACAGCTTCGACGTGAC 2195
      624 .....As 624
2196 GGGTCTGACAAATGTGTGCAAAAAACATTAACGAGATTAAGTGTG 2245
      :::::|
      624 nglYpHeValHlsLeuLeuAlaGluAsnIleLysLeuAspAsn..... 638
2246 CTTCATTGACTAAGACCGACATCAGCGGCAATGTCAT..... 2283
      639 .....SerLysValAspIleThrPheAspLysAspAsnSerGlnAsp 652
2284 ...CTTGCCGATTCAGCTCATTTAAATCTCACAGG...CTTGCCGACT 2327
      |||||
      653 ThrLeuAlaGlnThrAsnAlaGlnGlyMetAsnGlyLysValSerMetIle 669
2328 CAAGCGCAATCTTAGTCAAAATGCGCATACGCTTATACAGTCAGCCACA 2377
      :::::|
      669 eAsnSerHisIleLysIleValGlyAspGlyLysGlyLysGlyLysSerPro 686
2378 ACGGCACCCCAAAAGCGCAACCTT.....AGCCTGCTGGCAATGCC 2418
      :::::|
      686 hrgIyThrYrAlaThrMetPheLeuIleGlyLeuLeuIleGlyLys 702
2419 CAAGCAACATTTAATCAAGC.....ACATTAACGG 2450
      :::::|
      703 SerSerIlePheValLysSerHisGlnGlyYrThrPheLysThrAspPol 719
2451 CAACACATGCGCTTCGCGGC...AATGCTTCATTTATCTAAGCAGCACCA 2497
      :::::|
      719 yAspThrLysIleAlaGlyLysAsnSerLysGlyLysLeuLysIleThrA 736
2498 CCGTACAAAACGCGACGCTGACGCTTCGCGCAACGCTTAAGCAACCTA 2547
      :::::|
      736 latLeuSlnThrGly.....GlyArgAlaAlaGluGln 747
2548 AGGCATTCGCGACATCAGCGTATGTCTCCCTAAGCCGATTAAGCAATTT 2597
      :::::|
      748 .....LeuIleAsnGlyAlaLeuGlySerAlaAspAsnAspAla.. 760
2598 CCATTTTGAAGACGCGCTTACCGGCAATACAGCGGCGGCAAGGATA 2647
      :::::|
      761 AsnIleAlaAsnMetAlaPheThrIleGlyAspSerAlaAsnThrLysT 777
2648 CGGCATTACACTTAAAGACAGCAAGATGACGCTGCCGTCAGCGACGAA 2697
      :::::|
      777 hTrhIleGluAsnAlaAspIleThrAlaLeuAlaProAsnGly..... 791
2698 TTAGGCAATTTAAACCTTGACACAGCCACCTTACACTC..... 2736
      :::::|
      792 ...GlyThrAlaIyTrLeuSerSerLysAspAlaGlnIleGluValLysPr 807
2737 AATTCCGCGCATGCGCACGATGCGGAGGGGCGCAAAACGCGAGTGGCA 2785
      |||||
      807 oAsnSerAspPhe.....Thrphe 814
2786 CAGATGGCGCGCGCGCGCTTCGCGCGCTTCGCGCGCTTCCTATTATCC 2835
      :::::|
      814 hrgIleuLeuProArg..... 818
2836 GTTACACCGCAACTTGGTAGATCCGTTTCAACACGCTGACGGTAA 2885
      |||||
      819 .....GlyLysAsnLeuAsnGlnThrLysIleAs 828
2886 CGGC.....AAATGAACCGTACAGGAAACATTCGCTTATGTCGG 2926
      |||||
      828 nglYAlaSerThrLysLeuSerGlnAlaGlyPheAlaArgLeuYrAspL 845
2927 AACTCTTGCGCTACGCGACGCAAAATTAAGCTGCGGAAAGTTCGAA 2976
      RC
```

```
seq_name: SwissProt_40:PM20_CHLPN
seq_documentation_block:
ID PM20_CHLPN STANDARD: PRT; 1723 AA.
AC Q9Z812; Q9K2C1; Q9RB59;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable outer membrane protein pmp20 precursor (Polymorphic membrane
DE protein 20).
OS PMP20 OR CPN0540 OR CP0212.
GN Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.,
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey B.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., Deboy R., Kolony J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shida T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA";
RL Nucleic Acids Res. 28:2311-2314(2000).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
```


RX MEDLINE=202998986; PubMed=10833753;
 RA Shirai M., Hirakawa H., Onuchi K., Tabuchi M., Kishi F., Kimoto M.,
 RA Tateuchi H., Nishida J., Furukawa S., Fujinaga R., Yoneda H.,
 RA Matsushima H., Tanaka C., Tanaka K., Miura K., Nakazawa A.,
 RA Ishii K., Shiba T., Hattori M., Kuhara S., Nakazawa T.;
 RA Comparison of outer membrane protein genes omp and pmp in the whole
 RT genome sequences of Chlamydia pneumoniae isolates from Japan and the
 RT United States.;
 RL J. Infect. Dis. 181 Suppl 3:S524-S527(2000).
 CC -1- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.

DR	EMBL: AE001638; AAD18680.1; ALT_INIT.	
DR	EMBL: AE002181; AAF38082.1; -	
DR	EMBL: AP002547; BAA98746.1; -	
DR	EMBL: AB033817; BAA85968.1; -	
DR	PICT-2DPAGE; Q92812; -	
DR	TIGR: CP0212; -	
DR	InterPro: IPR003368; DUF145.	
DR	InterPro: IPR003571; OMF.	
DR	Pfam: PF02415; DUF145; 1.	
DR	Pfam: PF02385; OMF; 1.	
KW	Outer membrane; Signal; Multigene family; Complete proteome.	
FT	SIGNAL	21
FT	CHAIN	2
FT	DOMAIN	22 1723
FT	DOMAIN	94 97
FT	DOMAIN	140 144
FT	DOMAIN	147 160
FT	CONFLICT	1134 1134
FO	SEQUENCE	1723 AA; 179554 MM; 45075DBA0E3EE01 CRC64;

```
alignment_scores:      length: 1161
      quality:      195.00
      ratio:      0.400
      percent_similarity: 42.033
      percent_identity: 21.447
```

```
alignment_block:
US-09-303-518D-649 x PM20_CHLPN ..
```

Align seg 1/1 to: PM20_CHLPN from: 1 to: 1723

[illegible]

```

502 ..gIyAlAlIePhaIaValGlyGlyIleThrLeuSerAspLeuValGlu 517
970 CCACGTCGCAAAATGGGAATACTCTTTTAACGACGATAAATATGGCACAGG 1019
    ||| :||| :||| :||| :||| :|||
518 ValArg.....PheSerIysAsnIlysthrGlyAsnIlyrSerAla 530
    ..... :||| :||| :||| :||| :||| :|||
1020 AAAAATCATAGCCCAACATGACACACAACTTCGCTCATAGATTAAATAA 1069
    ||| :||| :||| :||| :||| :|||
530 aProIle...ThrIysAlaIaSerAsnThrAlaPro.....ValValS 544
1070 CACGACCGCTTCATTCGTTTAAATGTTTCTTATTCGACGACGACAGAGAA 1119
    :||| :||| :||| :||| :||| :|||
544 eSerSerThrThrAlaIaIaSerProAlaValProAlaIaIaIaIaIa 560
1120 CCTGTTATCATGCTGCAGAGTGGCTGCACACGATTAATGACCCAGA..... 1164
    ||||| :||||| ||| ||| :||| :|||
561 ProValThrAsnAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 577
1165 .....CTGATTAATGGAGAAATATTCCCTTTA 1192
577 uThrValSerGlyIleThrIserIleLeuSerPheGluAsnAsnGluCysG 594
1193 TTGACGACGAGAAAGGC..... 1209
594 IAsnAsnGlyGlyGlyIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 610
1210 .....GAATGATACCTTACCAGACAC.....ATCAATCAGCTGCTGG 1247
    :||| :||| :||| :||| :||| :|||
611 SerHisArgLeuGlnPheThrSerAsnIlyAlaIaIaIaIaIaIaIaIaIa 626
1248 AGGATTATATTCACAGACAGATTTACGTCGTG..... 1281
    ||||| ||| ||| :||| :||| :||| :|||
626 yGlyLeuIlyrCysGlyAspAspValaThrLeuThrAsnLeuThrGlyIysT 643
1282 .....CCTGAATAATACGAAACTTGGCAGAGCGCGGCGCTTATATC 1323
    ||||| :||| :||| :||| :||| :|||
643 hIleuPheGlnGluAsnSerSerGluIySHIsgIyGlyIleuSerLeu 659
1324 AGTGAAGACAGTACCGTTACTTGG..... 1347
    :||| :||| :||| :||| :||| :|||
660 AlaserGlySerLeuThrIleThrSerLeuGlnuserPheCysLeuAs 676
1348 .....AAAGTAAAGGC..... 1359
    ||| ||| ||| ||| ||| |||
676 nAlaSerThrAlaIaIaGluAsnGlyGlyAlaIaIaIaIaIaIaIaIaIa 693
1359 ..... 1359
693 IeValIeThrPheThrIlyrThrProThrProAsnGlnProAlaProVal 709
1360 .....GTGGCAAAAGCACCGCTGTGCCAA 1382
    ||||| :||| :||| :||| :||| :|||
710 GlnGlnProValIlyrGlyGluAlaIaIaIaIaIaIaIaIaIaIaIaIaIa 726
1383 AATGGCAAAAGCAGCGCTGCAGCTTCAACGCCAAAGGGCAAAAGCAAGCT 1432
    ||| ||| ||| :||| :||| :||| :||| :|||
726 sSerGlyIyGlyIleThrIlyrThrIysAsnAlaIaIaPheSerAsnLeuSerS 743
1433 CGATC.....AGCGTGGCGACGGTACAGTCATTTTG 1464
    :||| :||| :||| :||| :||| :|||
743 eIValIlyThrPheAspGlnAsnThrSerSerGluAsnIyGlyAlaIaIaIe 759
1465 GATTCAGCAGCGACGATTAAGGCAAAAGCAAGCCTTTAGTGAATTCGG 1514
    ||||| :||| :||| :||| :||| :|||
760 ThrGlnIyAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 774
1515 CTTTGTCAGCGGACGGGTACGGTGCACACGATGGCATATTCAGTTCA 1566
    :||| :||| :||| :||| :||| :|||
775 .....ThraValAsnIleThrAsn..... 781
1565 ACCCGCAACAACTGATTTGGCTTTGGCGGCGGCTTTGATTTAAAC 1614
    :||| :||| :||| :||| :||| :|||
782 .....AsnThrAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 792

```

1615 GGCATTCGCTTTCCTCCACCGTATTAACGATGAGGGCGCAT 1664
1693 GlycyllysalaHisphenaspargileasphen Le 804
1665 GATTGTCACCCCAATCAACAGACAA 1689
804 utrhvalginserasnlnalalalyslsglyglyValTyrleuGUA 821
1690 GAATCCACC 1698
821 spAlaleuileuGluValIleThrnglySerValSerGlnasnThr 837
1699 GTTACCATTCAGCCAAAT AAGATATT GC 1727
838 AlathrGlnserGlyglyTyrleTyrAlalysaspIleGlnleuGlnAl 854
1728 TACAAACCGCATACACAGCCTTGATGACCAAAAGAAATTCCTTACA 1777
854 aleuproglyserphenrIleThrAspAsnlyValGluThrSerleu. 870
1778 ACGGTTGGTTGGCAGAAAGATACGACCAAAACGAGGGGGGCTCAC 1827
871 ThrThrSerThrAsn 875
1828 CTGTTTACAGCCCGCGGAGAGACCGCATCCTGCTTCTCCGGCGG 1877
876 LeuTyrGlyGly 880
1878 AACAAATTTAAACGCAACATCACGCAACAAAC GCGAAACGT 1921
880 YIleTyrSerSerGlyValValThrleuThrAsnIleSerGlyThrPheG 897
1922 TTTTACGCGGAGACCAACAGCCGCGCTACATATTAAACGACCAT 1971
897 YIleThrnglyAsnSerVal IleasnThrAla 907
1972 TGGTCCGAAAGAGGGCATTCCTCCGGG 2001
908 ThrSerGlnaspAlaspIleGlnGlyGlyIleTyrAlaThrThse 924
2002 GAATGCTGTGGGACAAACGATCGA 2026
924 rIleuSerIleasnGlnCysAsnThrProIleuPheSerAsnAsnSerA 941
2027 TCACCGCACATTTAAAGCGAAACCTTCAATTAAGGCGGACAGCG 2076
941 lAlaIaThrlysthrSerThrThrlyGlnIleAlaGlyGly...Ala 956
2077 GTGGTTCCCGCAATGTTGCCAAAGTGAAGCGCATTTGAGCAA 2126
957 IlePheSerAlaAlaValThr IleGlnAs 966
2127 TCAGCGCCACAGCATTTT GGTGTGCAACCGCATTAAGCC 2167
966 nasnserGlnProIleIlePheleuAsnAsnSerAlaIysSerGluAla 983
2168 ACACAATCTGTACA CGTTGCGACTGCGAGCGT 2199
983 hrThAlaIaThrAlaGlyAsnlyAspserCysGlyAlaIaIleAla 999
2200 CTGCAAAATTTGTGGAAGAAACCAATTAACGACGA 2234
1000 AlaAsnSerValThrleuThrAsnAsnProGluIleThrPheGlySclAs 1016
2235 TAAAGTG ATTGCTCATTTGACTAAGACCG 2263
1016 nTyrAlaGlnThrGlyAlaIleGlyCysIleAspLeuThrAsnGlyS 1033
2264 ACATCAGCGCAATGTCGATCTGCGGATCGCATTTAATCTGCACA 2313
1033 erProProAlGlyValSerIleAlaAspAsnGlySerValIleuPheGln 1049

2314 GGCCTTGCCACACTACGCGCAATCTTAAGTCAAAATGCCGATACGTTA 2363
1050 AspAsnSerIleAsnAsnArgIlyAlaIleTyrGlyGlnThrIleAs 1066
2364 TACAGTCAACCCACAAACGCCCAAAACGCGCAAC 2397
1066 pIleSerArgThrGlyAlaThrPheIleGlyAsnSerSerlyHisAspG 1083
2398 CTTAAGCTCTGTGGCAATGCCAA 2421
1083 lySerAlaIleCysCysSerThrAlaIleuThrAlaProAsnSerGln 1099
2422 GCACAAAT ATCAAGCCACATTAACGCGCAACACATCGCTTC 2465
1100 leuIlePheGlnAsnAsnlyValThr GluThrThrAlaThr 1113
2466 GGGCAATGCTCATTTAAT 2484
1113 rThrlyAlaSerIleAsnleuGlyAlaAlaIleTyrGlyAsnAsnG 1130
2485 ..CTAAGCGACCGCGTA CAAACGCGAGTGTACG 2520
1130 IuThrSerAspIleThrIleSerleuSerAlaGlnAsnGlySerIlePhe 1146
2521 CTTTCCGCGAAC...GCTAAGCGCAACGTAAGCATTTCCGACATCAACG 2567
1147 PheIysAsnAsnleuGlyThrAlaThrAsnlyTyrCysSerIleAlaG 1163
2568 TAAATGTCCTTA GCCGTAAGGACATTTCC 2599
1163 yAsnValIlyPheThrAlaIleGlnAlaSerAlaGlyAlaIleSerP 1180
2600 ATTTGAACGACCGCCTTACCGGACAAATCAGCGCGGCAAGATGCG 2649
1180 heTyrAspAlaVal AsnValSerThrlyScluThrAsn 1192
2650 GCATTAACACTTAAGACAGCGAATGAGCGCTGCGTACGACGCAAT 2699
1193 AlaGlnGlnleuIlyAsnGlnIulysAlaThrSerThrGlyThrIle 1209
2700 A.....GSCAATTTAAACCTTGACACAGCGCACATTTACATCAATTCG 2743
1209 uPheSerGlyGlnleuHisGlnAsnlySerThrIleProGlnlyValT 1226
2744 CCGATGCGAC GATGCGGACGCGGCGCAAC 2775
1226 hrPheAlaHisGlyAsnleuIleleuGlyIysAsnAlaGlnleuSerVal 1242
2776 GCGAGTGCAGACAGATCGCGCGCGCGCTGCGCGCGCTGCGCGCTG 2825
1243 ValSerPheThrGlnSerProGlyThrThrIleThrMetGlyProGlyse 1259
2826 CCTATTATCG 2835
1259 rValleuSerAsnHisSerIlyScluaIaGlyIleAlaIleAsnAsnV 1276
2836 GTT 2838
1276 alIleIleAspPheSerGlnIleValProThrlyAspAsnAlaThrVal 1292
2839 ACACCGCAACT TCGGTAGAAATCCGTTTACACAGCGTGACGTAA 2885
1293 AlaProProThrleuIlySclValSerArgThrAsn...AlaAspSerIlyS 1308
2886 CGGCAAAATTAAGCGTCAAGGACATTCGCTTATGTGGAACCTCTTCG 2935
1309 AspIlyIleAspIleThrnglyThrValIleThrleu 1319
2936 GCATCCGCGACGCAAAATTTGAAGCTGCGGAAAGTTCCGAAAGCACTTAC 2985
1320 LeuAspProAsnGlyAsnleuTyrGlnAsnSerT 1331
2986 ACCTTGCGGTCACAAATACCGGCAACGAACTGCAAGCTCTGAAACAATT 3035

```

1331 ylleucly.....GluAspArgAspIleThrLeuPheAsnIle 1343
3036 GACGGTAGTGGAGAGAAAGAACAAACCGGTCTCGAAACCTTAATT 3085
1344 AspAsnSerAla.....SerGlyAlaValThrAla..ThrAsnV 1356
3086 TCACCGCTGCAAAACGAAACAGTCGATGCC.....GGCCCG 3120
1356 althreugngly...AsnleuglyAlaLysGlyTyrLeucllythr 1371
3121 TGG..... 3123
1372 TrpAsnLeuAspProAsnSerSerGlySerlyllelleleuLysTrpTh 1388
3124 .....CGTTACCACTCATCCGCAAAAGCGGAGCTTC 3157
1388 rPheAspLysTyrLeuArgTrpProTyrIleProArgAspAsnIlePheT 1405
3158 GCGTCGATTAAT.....CCGGTCAAAGAA 3180
1405 yllleAsnSerIleTrpGlyAlaGlnAsnSerLeuValThrValLysGln 1421
3181 CAAGAGCTTTCGCAAACTCGCGACGACGAAACCAAAACAGAGCG.. 3228
1422 GlyIleLeuGlyAsnMetLeuAsnAlaArgPheGlnAspProAlaTh 1438
3229 .....GAAAGACAGACG 3241
1438 eAsnAsnPheTrpAlaSerAlaIleGlySerPheLeuArgLysGluValS 1455
3242 GCGAAGCCTTGACGCGCTGATTCGGCGCGCGC..... 3276
1455 eArgAsnSerAspSerPheThrGlyArgGlyTyrThrAlaAla 1471
3277 ...GATGCCGTGCAAAACAGCAAGAGCGTT 3303
1472 ValAspAlaLysProArgGlnIlePheIle 1481
seq_name: SwissProt_40:VAC4_HELPY
seq_documentation_block:
ID VAC4_HELPY STANDARD; PRT; 1291 AA.
AC Q48258;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Vacuolating cytotoxin precursor.
GN Helicobacter pylori (Campylobacter pylori).
OS Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=185-44;
RX MEDLINE=94335650; PubMed=8057855;
RA Haas R., Schmitt W.;
RT "Genetic analysis of the Helicobacter pylori vacuolating cytotoxin:
RT structural similarities with the Iga protease type of exported
RT protein."
RT Mol. Microbiol. 12:307-319(1994).
RU Mol. FUNCTION: INDUCES VACUOLATION OF EUKARYOTIC CELLS. CAUSES
CC -1- ULCERATION AND GASTRIC LESIONS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sdb.ch).

```

```

CC -----
DR EMBL; Z26883; CA81528.1; -
DR InterPro; IPR003842; Vaca.
DR Pfam; PF02691; Vaca; 1.
KW Cytotoxin; Toxin; Signal.
FT SIGNAL 1 33
FT CHAIN 34 2
FT PROPEP 2 1291
FT SEQUENCE 1291 AA; 139635 MW; ECA56A61CAEF3669 CRC64;

```

```

alignment_scores:
  Quality: 194.50      Length: 1396
  Ratio: 0.331         Gaps: 66
Percent Similarity: 42.049      Percent Identity: 18.983

```

alignment_block:

US-09-303-518d-649 x VAC4_HELPY ..

```

Align seg 1/1 to: VAC4_HELPY from: 1 to: 1291
595 GCAAGCGACGCAATATGGCGATCTGAT.....GANGA 626
118 AlaIaArgHisTyrTrpValLysAspGlyIntPrAsnLysLeuGluVa 134
627 TGACCCCAATACCCGCAAAAGTTCATATCATATGCAAGTGGTATCTT 676
134 AspMetGlnAsnAlaValGlyThrTyrAsnLeuSerGlyLeuIleAsn 151
677 GCGTCGTTGTT..... 687
151 heTrnGlyGlyAspLeuAspValAsnMetGlnLysAlaThrLeuArgLeu 167
688 .....GGCAATACCTTT..... 699
168 GlyGlnPheAsnGlyAsnSerPheThrSerTyrLysAspSerAlaAsp 184
700 .....GCACAAAT..... 708
184 gThrThrArgValAspPheAsnAlaLysAsnIleLeuIleAspAsnPhel 201
709 .....GGATCAGGTGGTGGCAGACAGTCACACTTATGTT 738
201 euGlnIleAsnAsnArgValGlySerGlyAlaGly..... 212
739 AGTGAAAATTAATTAACATAGCCCATATGTTTATACCAAGAGAGCTC 788
213 ...ArgLysAlaSerSerThrValLeuThrLeuGlnAlaSerGluGlyI 228
789 ATTTGGCAGACAGTGGCTCACCAATGTTATCTATGATGCCCAAAAGCAA 838
228 eThrSerArgGlnAsnAlaGluIleSerLeuTyrAspGlyAlaThrLeu 245
839 AGTGTTAATTAATGAGGTATTCGAACGCGCAACCCCATATATAGAA 888
245 snLeuAlaSerAsnSerValLysLeuMetGlyAsnValTrpMetLysArg 261
889 AGCAAT.....GGCTCCAGCTG..... 906
262 LeuGlnIlyValGlyAlaLysLeuAlaProSerTyrSerThrIleAsnTh 278
907 .....GTTCTTAAGATTGGTTCTATGATGAATCTTGGCTGAGAGT. 948
278 rSerLysValThrGlyGluValAsnPheAsnHisLeuThrValGlyAspH 295
949 .....ACCCAT..... 954
295 isAsnAlaAlaGlnAlaGlyIlelleAlaSerAsnLysThrHisIleGly 311
955 .....TCAGATTCTACCAACACAG 974
312 ThrLeuAspLeuTrpGlnSerAlaGlyLeuAsnIlelleAlaProIle 328

```

[illegible]

```

838 yAsnThrThrAsnLeuProThrAsnThrAsn...LysValArgPhe 854
2582 CGGATAGGAGGATATCCATTGGAAGCAGCCGCTTACCGACAAATC 2631
854 IleValArgPhe...IleValArgPhe...LysValArgPhe 868
2632 AGCGGGGCGGAGGATGCGGATTTACCTTAAAGACAGGATGAGCGT 2681
869 AlaThrProAsnLeuValAlaIleAsnGlnHisAspPhe...GlyThr 884
2682 GCGGTGAGGAGGATGCGGATTTAACTTGAACAGGCGGACCATTA 2731
884 eGluSerValPheGluLeuAlaIleAsnArgSerAsnAspIleAspThrLeu 901
2732 CACTCAATTCGCGCTATCGC... 2751
901 yAlaAsnSerGlyAlaGlnGlyArgAspLeuGlnThrLeuLeu 917
2752 ...CAGCATGCGGAGGCGGCGCAACCGGACGATGCGACAGATCGCC 2795
918 AspSerHisAspAlaGlyTyrAla... 925
2796 GCGCGCGCGTTCGCGCGCTTCGCGCTTCCTATATTCGCTTACACCGC 2845
926 ...ArgThrMetIleAspAlaThrSera 934
2846 CACTTCGCTGATATCCGCTTCAACAGCGTGAACGCGCAATTCG 2895
934 IaAsnGlnIleThrLysGlnLeuAsnThrAlaThrThr 948
2896 AACGCTGAGGAGAACATTCGCTTATGTCGGAACCTTCGCGTACCGCAG 2945
949 AsnAsnIleAlaSer... 953
2946 CGACAAATTCGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACTTGGCG 2995
954 ...LeuGlnHisLysThrSerGlyLeuGlnThrLeuSerL 966
2996 TCACAAATTCGCGCAACGAACTGCAAGCTTCGAACAAATTCGAGTACGT 3045
966 euserSn... 968
3046 GAAGGAAAGACACAAACCGCTGTCGGAACCTTAAATTCACCTGCA 3095
969 ...AlaMetIleLeuAsnSerArgLeuVal 977
3096 AAC...GACACGTCGATGCGGCGGCTTCGCTTACCTGCA 3127
977 IaAsnLeuSerArgArgHisThrAsnHisIleAspSerPheAlaLysArgL 994
3128 ACCCACTCATCCGCAAAAGACGCGGAGTTCGCGCTGATATCCGCTGCA 3177
994 euglnAlaLeu...Lys 998
3178 GAACAGAGCTTTCGCAAACTCGGCAAGGAGAA... 3213
999 AspGlnArgPheAlaSerLeuGlnSerAlaGlnValLeuTyrGlnPhe 1015
3214 GCGCAAAACAGCGGCAAAAGACAGCGGCGGACGCTGAGCGCTGA 3262
1015 eAlaProLysTyrGlnLysProThrAsnValTyrPhe... 1027
3263 TTGCGCGCGGCGGATCGCTCGAAAGACAGAAAGCGTTCGCGCAACG 3312
1028 ...AsnAlaIleGlyGlyThr 1033
3313 GCGCGGAGGAGCGCGGAAATGTCGCTTATGACAGGCGGAGAGAGA 3362
1034 SerLeuAsnSerGlyGly...AsnAlaSerLeuTyr... 1044
3363 GAAAAACGCGTGCAGCGGATTAAGACACGCTTCGCGAAACGCGCG 3412
1045 ...GlyThrSerAlaGlyValAlaAspAlaTyrLeuAsnGlnGlyValG 1059

```

```

3413 AACGGAAACCCGGCGCTACACCGGCTTCCCCGGCGCGCGCGCC 3462
1059 lAla... 1060
3463 CGCGGGGATTTGCGCAACTGCACCCGACCGGACGCCCAACGCGAGC 3512
1060 ... 1060
3513 CGACGTATCAGCCGTTATGCCAATAGCGCTTGTAGTAAATTTCCGCA 3562
1061 ...IleValGlyGlyPheGlySerTyrGlySerPheSerAsnG 1076
3563 CGCTCAACAGC... 3573
1076 lAlaAsnSerLeuAsnSerGlyAlaAsnAsnThrAsnPheGlyValTyr 1092
3574 ...GTTTCGCGCTACAGGACGAATTAAGC...CGCGT 3605
1093 SerArgIlePheAlaAsnGlnHisGlnPheAspPheGlnAlaGlnGlyAl 1109
3606 ATTGCGGAGAGCGCGGCAACCGCTTGTGACACAGCGC...ATCCGGG 3652
1109 aLeuGlySerAspGlnSerSerLeuAsnPheLysSerAlaLeuValArg 1126
3653 AC...ACCAACAC 3663
1126 sPheAsnGlnSerTyrAsnTyrLeuAlaTyrSerAlaAlaThrArgAla 1142
3664 TACCGTTGCAAGATTTCCGCGCTACCGCAACAAACGACCTGGCGCA 3713
1143 SerTyrGlyTyrAspPheAlaPhePheArgAlaLeuValLeuLysP 1159
3714 ATTCGATGCAAAACCTCGGCGAGCGCGCTGCGATCTGCTT 3763
1159 O...SerValGlyValSerTyr 1166
3764 CGCACACCGGAGCAAAACCTTCGACGAGCGCATCGCACTCGCA 3813
1166 snHisLeuGlySerThrAsn...PheLysSerAsnSerAlaGlnVal 1181
3814 CGGCTGCGCGCGCGCGCTTTCGCGCAATGCGCATCGACAGTCTCA 3863
1182 AlaLeuLysAlaGlnGlyAla... 1187
3864 CATCGCATCAGCGCGCGGCTTTCGACGAGCGGACCTTCAGAGC 3913
1188 ...SerSerGlnHisLeuPheAsn 1195
3914 GCATCGGAGCAAAATCCGCGCGGCTGCTGATTCAGGCTTACGCA 3963
1195 lAsnAla...AsnValGlnAla 1201
3964 CGATACCGCGCGGCTTTCGCGGATTCGCGCATGCAACCGCATCGCGC 4013
1202 ArgTyr...TyrTyrGlyAs 1207
4014 AACCGCTATTCGTCGCAAAAGCGGATTAACGCTTCGAAACGTCATA 4063
1207 pThrSerTyrPheTyrMetAsnAlaGlyValLeuGlnGlnPheAlaAsn 1224
4064 TC...GCCACCGCGCGCTTCGATTCACACCGCTTACCGC 4098
1224 heGlySerSerAsnAlaValSerLeuAsnThrPheLys 1236

```

seq_name: SwissProt_40:PM2L_CHLPPN

seq_documentation_block:
 ID PM2L_CHLPPN STANDARD; PRT; 1609 AA.
 AC Q92605; Q9RB58;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

DE Probable outer membrane protein pmp21 precursor (Polymorphic membrane
GN protein 21).
GN PMP21 OR CPN0963 OR CP0897.
OC Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RX NCBI_TaxId=83558;
RX (1)
RM SEQUENCE FROM N.A.
RM STRAIN-CWL029.
RC MEDLINE=99206606; PubMed=10193388;
RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.,
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.",
RT Nat. Genet. 21:385-389(1999).
RM (2)
RM SEQUENCE FROM N.A.
RM STRAIN-AR39.
RC MEDLINE=20150255; PubMed=10644935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Uppertack T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dolson R.,
RA Gwyn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eelsen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.",
RT Nucleic Acids Res. 28:1397-1406(2000).
RL [3]
RP SEQUENCE FROM N.A.
RC STRAIN-J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kihara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.",
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -1- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001676; AAD19099.1; -
CC EMBL; AE002248; AAF38684.1; -
CC EMBL; AP002548; BAA99171.1; -
CC PHC1-2DPAGE; Q9RB58; -
CC TIGR; CP0897; -
CC InterPro: IPR003357; OMP.
CC Pfam: PF02385; OMP; 1.
KW Outer membrane; Signal; Multi-gene family; Complete proteome.
KW SIGNAL 1 30
KW CHAIN 31 1609 POTENTIAL.
FT CONFLICT 420 420 I->M (in ref. 3).
SQ SEQUENCE 1609 AA; 170865 MW; 2604C3E91CD4024CB CRC64;

```

alignment_scores:	
Quality:	194.50
Ratio:	0.332
Percent Similarity:	41.531
	Length: 14111
	Gaps: 76
Percent Identity:	20.051

alignment_block:
US-09-303-518D-

Align seg 1/1 to: PM21_CHLPN from: 1 to: 1609

1108 ACAGCAGA.....GAACCTGTTATCATGCTGCAGGTGTCACACAG 1151
||||||| ||| ||| ||| |||

1152 TTAATGACCCACACACTGAATATGAGAAATAATTCCTTATTAGCAGAC 1201
1340 T.....ValValGly 344
1202 GAAGAAGCGAATGATTACTTCCAGCAC...ATCAATCAAGGTGCTGGA 1248
344 snlysgcllyalailevalalevalalelulysnsderalaglylserasnly 360
1249 GGATTATATTTCCAGAGAGATTTTACGCTGTCGCCCTGAATAATACGAAC 1296
361 glyalaphaalacysglyserphevaltylser.....AsnAsnGlu... 374
1299 TTGGCAAGCGCGCGGCTCATATCACTAGTAAGACACTACCGTTACTTGGA 1348
375AsnThrAlaLeuTrpL 380
1349 AAGTAAAC.....GGCGTGCAAAAGCAGCGCGCTGCCAAATC 1386
380 ysgluasnnglnalaleuSERglYglYalaleSerSerAlaSerAspIle 396
1387 GGC...AAAGCAGCGGTGCACGTTCAAGCCAAAGGGAACCAAGGCTC 1433
397 AsplleclnclYasnCySserAlaIleGlupheserGlyasnGlnSerIe 413
1434 GATAGCGGTGGCGGACGGTCAAGTCATTTTGGATCAGCAGACGACGATA 1483
413 uilelalelenglYglunhlsIleGlYleuThrAsp..... 424
1484 AAGCAAAAAACACAGCTTACTGAAATTCGCTTGAGCGGACGGGT 1533
425PheValGlyglYglYalaleuAlaAlaGlnGly 435
1534 ACGGTCAACTG.....AATCGCATATCAAGTTC.....AACCCGA 1571
436 ThrleuThrleuAlaGAsnAsnAlaValalGlnCyValLysasnThrSe 452
1572 CAACACTGATTTTGGC...TTTGGCGGCGACGTTTGATTTAACGGCG 1618
452 TlySThrHISglYglYalaleleuAlaGlyTThrValSpleAsn... 467
1619 ATTGCGTTTCGTTCCACCGGTATTCAAAATACCGATGAGGGCGATGATT 1668
468GluThrIleSerGluValAlaPhe... 475
1669 GTCAACCAACATCAAGACAAGATTCACCGGTATCCATTTCAGCGATTA 1718
476LysGlnAsnThrAlaAlaLeuThrclYglY... 485
1719 ACATATTCGTCACACCGGCAATAACAACAGCTTGATAGCAAAAAAGAA 1768
486AlaLeuSerAlaAsnAsp.....LysValIleI 495
1769 TTGCCTACACAGGTGGTTGGTTGGC.....GAGAAACATACGACC 1806
495 lealaasnAsn.....PheclYglIleLeuPheIleuGlnAsnGluVal 509
1807 AAAACGAACGGCGCGCTCAACCTGTTTAC.....CAGCC 1841
510 ArgasnHISglY.....GlyAlaIleIleTyrcysglYcysArgSerAsnPr 524
1842 CGCGCGAAGACCGCACCTGTGCTGTTCGGGGGAACCAATTAAAC 1890
524 olyleuEnglunhlyS.....AspserIleGluAsnIleAsnI 537
1891GCCAACATCCAGCAAAACAAACGCGCAACTGTT... 1923
537 lelleeglyasnSerGlyAlaIleThrPheLeuYasnlyAlaSerVal 553
1924TTCAAGGCGACACCAACACCGCA 1946
554 leuglYvalMetThrGlnIleGluAspYrlyAlaGlyglYglYalaleuTr 570

RA Nies H., Nissen-Meyer J.;
 RT "Purification and N-terminal amino acid sequence determination of the
 RT cell-wall-bound proteinase from *Lactobacillus paracasei* subsp.
 RT *paracasei*." *J. Gen. Microbiol.* 138:313-318(1992).
 RL J. Gen. Microbiol. 138:313-318(1992).
 CC -1- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
 CC GROWTH OF THE BACTERIA ON MILK.
 CC -1- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
 CC specificity, although some substrate preference have been noted,
 CC e.g. large hydrophobic residues in the P1 and P4 positions, and
 CC Pro in the P2 position. Best known for its action on caseins,
 CC although it has been shown to hydrolyse hemoglobin and oxidized
 CC Insulin B-chain.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 CC SUBTILASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M83946; AAA25248.1; -;
 DR PIR; B44858; B44858.
 DR HSSP; P00782; 1S01.
 DR MEROPS; S08.019; -;
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR003137; PA.
 DR InterPro; IPR002029; Peptidase_S8.
 DR Pfam; PF00746; Gram_pos_anchor; 1.
 DR Pfam; PF02225; PA; 1.
 DR Pfam; PF00082; Peptidase_S8; 3.
 DR PRINTS; PR00723; SUBTILASIN.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SRS; 1.
 DR PROSITE; PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
 DR HydroLase; Serine protease; Cell wall; Zymogen; Signal;
 KW Transmembrane.
 KM
 FT SIGNAL 1 33
 FT PROPEP 34 187
 FT CHAIN 188 1876
 FT DOMAIN 1877 1895
 FT TRANSMEM 1896 1902
 FT DOMAIN 1896 1902
 FT ACT_SITE 217 217
 FT ACT_SITE 281 281
 FT ACT_SITE 620 620
 FT DOMAIN 1867 1872
 SO SEQUENCE 1902 AA; 200253 MW; D8C9F38CE5DA582 CRC64;
 alignment_scores: Length: 1327
 Quality: 194.50 Gaps: 65
 Ratio: 0.345 Percent Identity: 19.819
 Percent Similarity: 42.502
 alignment_block:
 US-09-303-518D-649 x P2P_LACPA ..
 Align seg 1/1 to: P2P_LACPA from: 1 to: 1902
 415 AAAATGTGAACGCAATATTAAGACGAGCTAAAGCCATCTT 464
 836 LysIleValaspSerLeuasn.....GlyIleThrTyIserProal 849
 465 TGGCGCGCGATTAT...CATATGCCCGCTTG.....C 493
 849 agIlyIasnTyIglYthrValProLeuThrAsnLysAsnThrIglY 866

494 ATAAATTT.....GTCCAGATGACAGAACCTGTGAATGACC 531
 866 IsgIntYrTyIglYglYmetValThrAspAla..... 876
 532 AGTATATGATGATGGCGGAATATATGATCAATAATATACCCGACCG 581
 877AspGlyLysGlnThrValaspAspGlnAlaIleAla..... 888
 562 TGTTCGATTGGGCGACGACGACGCAATTTGGCGATCGATGAAGATGAC 631
 889PheSerSerasp..... 892
 632 CCAATACCGCGAAGATTCATATCATATTCAGATGCGATTTCTGCGTC 681
 893LysAsnAlaLeuThrAsnAspIleSerMetGlnTyIleu 906
 642 GTTGTGCAATACCTTTGACAA.....AATGATCAGGTG 719
 907 LeuATGAsnIleSerAsnValGlnValaspIleLeuAspGlnGlnLys 923
 720 TGGCAGATGCAACTAGTAGTGAATAAATTAACATAGCCCATATGTT 769
 923 nlyValThrThrLeuSerSerSerThrAsnGlnThrIleThrTyI 938
 770 TTTTACCAACAGAGAGCTCATTTGGCGACAGCTGCTACCAATGTTATC 819
 938 938
 820 TATGATGCCCAAAAGCAAAAGTGTATTAATGAGGATTCGCAAGCG 869
 939 TyrAspAlaHisSerGlnIleTyIle..... 947
 870 CAACCCCTATATAGAAAAAGCAATGCTTCACGCTGCTGTAAGATT 919
 947 947
 920 GGTTCATATGTAATCTTTGCTGAGATACCATTCAGTATTCAGAA 969
 948 ..TyrTyIasn.....AlaProIalThrAspGlyThrTyIasp 960
 970 CCAGTCGCAAAATGGAATATCTTTTACGACGATTAATATGCGACAG 1019
 961 GlnArg...AspLysAsnIleLysThrAlaaspAspGlySerTyIThrTy 976
 1020 AAAATCATGCTCAACATGACACACATTCCTGCTATATAGATTAATA 1069
 976 rArgIleSerGlyValProGlnGlyLysp..... 986
 1070 CAGAACCGTTCATATGTTATGTTTCTTATCCGACGACGACAGAA 1119
 987LysArgGlnValPheAspValProPheLysLeuAspSerLysAla 1001
 1120 CCT...GTTATCATGCTGCGAGGTGCTGTCACAGCTTATTCAGCCAGCT 1166
 1002 ProThrValArgHisValAlaLeuSerAlaLysThr..... 1013
 1167 GAATATGAGAAAAATTTCTTATTCGACAGGAAAGGCGAATGGA 1216
 1014GluAsnGlyLysThrGlnIleTyI 1021
 1217 TACTTACGACGACATCATCAAGGTGCTGAGAGATTAATATTCACAGA 1266
 1021 yIleThrAlaGlnAlaLysaspAspLeuSerGlyLeu..... 1033
 1267 GATTTACGCTGCTGCGCTGAAATATACGAACATTGGCAGGCGCGGCG 1316
 1034 AspAlaThrIleLysSer..... 1038
 1317 TCATATCATGTAAGACAGTACCTTACTGGAAGTAACGGCGTGACAA 1366
 1039ValLysThrAlaIleAsnGlnValThr 1048

1541 InAspValProAlaAlaGlyValThrValThrAlaAsnGlyThrPheLys 1557
2917 TTTATGTCG...GAACTCTTGCGCTACCGACGCAAAATGAACTGGC 2963
1558 PheLysSerThrAspLeuTyrGlyAsnGlySerProAlaValAspTyr.. 1573
2964 GGAAGTCCGAAAGGCACTTACACCTTGCGGCTCAACATACCGGCAAG 3013
1574ValValThrAsnIleLysAlaAspA 1582
3014 AACCTGCAGCGCTGACAAATTTGACGCTAGTGGAGAAAACAAACAA 3063
1582 sPrProAla.....GlnLeuGlnThrAla LysGlnAlaLeuThrAs 1595
3064 CCGCTGTCGAAACCTTAATTTACCCCTGCAAAAGCAACGCTGCAATG 3113
1595 n.....LeuIleAlaSerAlaLysThr.....LeuS 1604
3114 CGGCGCGTGGCGGTTACCACTCATCCGCAAAAGCGCGAGTTCGCGCTGC 3163
1604 erAlaSerGlyLysTyrAspAspAlaThrThrThrAlaLeuAlaAla... 1619
3164 ATMATCCGCTCAAGACACAGAGCTTTCGCAACACTCGCAAGGCGAGAA 3213
1620AlaThrGlnLys 1623
3214 GCCAAAAACAGCGCGGAAAAAGACACGCGCAAGCCTTGACGCGCTGAT 3263
1623 sAlaGlnThrAlaLeuAspGlnThrAsp..AlaSerValAspSerLeuTh 1639
3264 TCGCGCGCGGCGCGATGCCGTGCAAAAGCAAGAAAGCTTCCGCAACCG 3313
1639 rGlyAlaAsnArgAspLeuGlnThrAlaIleAsnGlnLeuAla..... 1653
3314 CCGCGAGCGAGCGGCGGAAAAATGTCGCAATATGACAGCGAGAGAGAG 3363
1653 1653
3364 AAAAAGGCTGCGAGCGGATAAAGACACCGCTTGCGCAACAGCGGA 3413
1654 ...AlaLysLeuProAlaAspLysLysThrSerLeuLeuAsnGlnLeuG 1669
3414 AGCG.....GAACCGCGCGCG 3430
1669 nSerValLysAlaAlaLeuGlyThrAspLeuGlyAsnGlnThrAspPro 1686
3431 CTACC...ACGCGCTTCCCGCGCGCGCGCGCGCGCGGATTTGCGCG 3477
1686 erThrGlyLysThrPheThrAlaAla.....LeuAsp 1696
3478 CAACCTGCACCCCAACCGCGAGCCCAACCGCGAGCGGACCTGATCAGCG 3527
1697 AspLeuValAlaGlnAlaGlnAlaGlyThrGlnThr..... 1708
3528 TTTATGCCAATAGCGTTGAGTGAATTTCCGCGACGCTCAACAGCGTTT 3577
1709AlaAspGlnLeuGlnAlaSerLeuAlaLysVal.. 1719
3578 TCGCGCTAGAGGACGAATTAGACCGCGTATTTGCGCGAG..... 3616
1720LeuAspAlaValLeuAla..LysLeuAlaGlnG 1730
3617 ...ACGCGCGCAACGCG.....TTTGCACAGCGCGCAT 3647
1730 ylleLysAlaAlaThrProAlaGlnValGlyAsnAlaLysAspAlaAla 1747
3648 CCGGG.....ACACCAAACTACCGTTCGCAAG 3676
1747 hrGlyLysThrTrrPyrAlaAspIleAlaAspThrLeuThrSerGlyGln 1763
3677 ATTTCGCGCGCTACGCGCAACCAACGACCTGCGCAATGATGCGAG 3726
1764 AlaSerAlaAspAlaSerAspLysLeuAlaHisLeuGlnAlaLeuGlnSe 1780

3727 AAAAACCCTGCGACGCGCGCGCTCGGCATCCGTTTTCGACAAACCGGAC 3776
1780 rLeuLysThrLysValAlaAlaValAlaGlnAlaAlaLysThrAlaGly.. 1796
3777 CGAAACACCTTCGACGAGCGCATCGGCAACTCGGCACGCG..... 3817
1797 ..LysGlyAspAspThrThrThrGlyThrSerAspLysGlyGlyGlnGly 1812
3817 3817
1813 ThrProAlaProAlaProGlyAspThrGlyLysAspLysGlyAspGlnG 1829
3818TTGCCACGCGCGCGCTTTTTCGCGCA 3842
1829 ySerGlnProSerSerGlyLysAsnIleProThrLysProAlaThrThrT 1846
3843 ATACGCGATCGACAGGTTCATCATCG 3868
1846 hrSerThrSerThrAspAspThrThr 1854